Anonymitet i statistikk

En casestudie av hvordan Statistisk sentralbyrå går fra å behandle entydig identifiserende personopplysninger til å publisere anonym statistikk

Johanne Hafnor

Masteroppgave ved Avdeling for forvaltningsinformatikk
Det juridisk fakultet

UNIVERSITETET I OSLO

10.05.2019

År: 2019

Tittel: Anonymitet i statistikk - En casestudie av hvordan Statistisk sentralbyrå går fra å behandle entydig identifiserende personopplysninger til å publisere anonym statistikk

Forfatter: Johanne Hafnor

http://www.duo.uio.no/

Sammendrag

Masteroppgaven tar for seg metoder og teknikker Statistisk sentralbyrå benytter for å anonymisere statistikken de utarbeider der opplysningene stammer fra et enkeltindivid.

Oppgaven starter med å forklare hva som ligger i begrepet personopplysning og hvordan ulike grader av identifisering har ulikt beskyttelsesbehov og lovlighet. Oppgaven beskriver videre flere teknikker som kan benyttes for å beskytte de identifiserende elementene. Hensikten med teknikkene er å støtte opp under personvernprinsippene for å etterleve krav i personvernforordningen.

Avslutningsvis beskrives hvordan Statistisk sentralbyrå behandler personopplysninger internt og hvordan de utarbeider statistikk der opplysninger stammer fra et enkeltindivid. Metodene SSB bruker skal sørge for at statistikkene de publiserer er anonyme. I oppgaven belyses både juridiske, informatiske og organisatoriske virkemidler, som byrået har tatt i bruk for å understøtte sitt ansvar både som statistikkmyndighet og som behandlingsansvarlig for personopplysningene til hele Norges befolkning.

Forord

Det siste året har vært et utfordrende men også et svært lærerikt år. Gjennom denne masteroppgaven har jeg tilegnet meg ny kunnskap om identifisering, beskyttelsesteknikker og Statistisk sentralbyrås praktisering av anonymisering i statistikk.

Jeg vil gi en stor takk til min veileder Dag Wiese Schartum for raske og konstruktive tilbakemeldinger igjennom hele prosessen. Jeg vil også takke mine to informanter fra Statistisk sentralbyrå som var svært behjelpelige og kunnskapsrike.

Til slutt vil takke mine gamle medstudenter for støttende ord det seneste året. Dere vet hvem dere er.

Innholdsfortegnelse

Innledning

1.1 Tema, bakgrunn og aktualitet

Anonymisering er viktigere og samtidig vanskeligere enn noen gang sies det. Ny teknologi gir nye muligheter for innsamling og bruk av informasjon om enkeltmennesker. I takt med et økende behov for å gjenbruke og analysere innsamlede data er anonymisering av personopplysninger sentralt for å utnytte verdien som ligger i dataanalyse på en personvernvennlig måte. Utnyttelse av data har i flere år blitt omtalt som `den nye oljen` og begrepet stordata blir brukt som en mulig løsning på store utfordringer samfunnet vårt står ovenfor. Når det kommer til stordata er det å lagre store datamengder praktisk uproblematisk da datateknologien er blitt stadig kraftigere og billigere. Det finnes ikke en klar definisjon på hva stordata er (fra engelske: Big Data), men artikkel 29-gruppen, som var EUs rådgivende organ på personvernspørsmål, har beskrevet stordata som gigantiske mengder digitale data som er kontrollert av selskap, myndigheter og andre organisasjoner, og som gjøres til gjenstand for omfattende analyse. Andre kilder viser til stordata som en idé om å få noe meningsfylt ut av all tilgjengelig data.

Samtidig som det teknologiske godstoget har satt opp farten i form av stadig kraftigere og billigere datateknologi har EU strammet inn på personvernregelverket. Den 25. mai 2018 ble EUs personvernforordning implementert i europeisk rett. Med det nye regelverket har det kommet en del nye plikter for den behandlingsansvarlige og styrkede rettigheter for den registrerte. Personvernlovgivningen i Norge har de siste 20 årene dog inneholdt mange av de samme kravene som forordningen tydeliggjør og konkretiserer. Uavhengig av endringene i regelverket for Norge sin del, har personvernforordningen fått stor oppmerksomhet både i og utenfor EU. Dette synliggjøres blant annet i mediedekningen her i Norge. Mangelfull etterlevelse av personopplysningsloven og brudd på personopplysningssikkerheten fører stadig til medieoppslag. I Helse Sør-Øst saken fra mai 2017 avslørte NRK at IT-arbeidere fra Asia og Øst-Europa har hatt tilgang til sensitive personopplysninger. I 2013 ble medlemslistene til Høyre publisert på Facebook. Bergen kommune har hatt datainnbrudd i administrasjonssystemet som har medført at uvedkommende har fått tilgang til personopplysninger. Felles for de tre nevnte eksemplene er at de behandlingsansvarlige har fått sterk kritikk for håndteringen av bruddene på personvernet og

personopplysningssikkerheten. Det at det har vært et brudd i seg selv har fått mindre fokus. Datatilsynet informerer samtidig i sin årsrapport for 2018 at antall avviksmeldinger[1] har mer enn tredoblet seg fra året før. Antall avvik indikerer neppe at sikkerheten har blitt betydelig svekket siden forrige år, det er nok heller den store oppmerksomheten personvernforordningen har hatt.

Årsaken til at stordata ikke ennå er fullt utnyttet skal jeg ikke forsøke å svare på i denne oppgaven. Isolert sett kan det være mange årsaker til at ikke full utnyttelse av dataanalyse er satt i gang. Det kan være at det juridiske feltet er krevende og ikke muliggjør denne typen tilgjengeliggjøring, eller at anonymiseringsteknikker er så utfordrende at den behandlingsansvarlige ikke tør å offentliggjøre data. Stordata gjør at anonymiseringen blir mer utfordrende enn tidligere da det er en større mulighet for reidentifisering. Anonymiseringsteknikkene som benyttes skal ikke bare sørge for at dataene er anonyme i dag, men også i fremtiden, i en verden der teknologien stadig utvikler seg.

Når det kommer til behandling av personopplysninger så koker dette ansvaret ned til en behandlingsansvarlig som er forpliktet til å behandle personopplysningene forsvarlig etter personopplysningsloven. Ved å studere en anonymiseringsprosess hos en behandlingsansvarlig som er pålagt å publisere statistikk, vil jeg se på de vurderingene som blir gjort knyttet opp mot anonymisering av personopplysninger og hvilke metoder og teknikker som benyttes. Jeg har valgt Statistisk sentralbyrå som case, som er det nasjonale organ for utarbeidelse av offisiell statistikk om Norge.

[1] Etter personvernforordningens artikkel 33 "Melding til tilsynsmyndigheten om brudd på personopplysningssikkerheten".

1.2 Forskningsspørsmål og avgrensning

For å kunne belyse temaet jeg har beskrevet i innledningen har jeg avgrenset meg til tre forskningsspørsmål som skal besvares.

1) Hvilke rettsregler gjelder for identifisering og for beskyttelse av identiteter knyttet til personopplysninger?

I forbindelse med forskningsspørsmål 1 vil jeg utføre en rettslig analyse av gjeldende rett innenfor personopplysningsregelverket. Her vil jeg gjennomgå personvernbegrepet generelt og se på ulike grader av identifisering.

2) Hvilke teknikker kan benyttes for å beskytte de identifiserende elementene?

I forskningsspørsmål 2 går jeg inn på hvordan personopplysninger kan beskyttes ved hjelp av forskjellige teknikker. Denne delen av oppgaven tar for seg personopplysningssikkerheten ved å beskrive to ytterpunkter av mulige identifiserende personopplysninger, fra det behandles entydige identifiserende personopplysninger til helt anonyme opplysninger. Beskyttelsesmekanismer innenfor de to ytterpunktene blir forklart og drøftet.

3) Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym?

Forskningsspørsmål 3 er spesielt relevant for casen, der jeg studerer hvordan Statistisk sentralbyrå (heretter kalt SSB) sørger for at statistikken de offentliggjør er anonym. Dette er hoveddelen av masteroppgaven hvor jeg har intervjuet to ansatte i SSB for å få svar på hvordan de praktiserer anonymitet i statistikk.

Ved å undersøke hvordan SSB anonymiserer statistikkene de publiserer ser jeg på hvordan SSB benytter juridiske, tekniske og organisatoriske virkemidler for å utføre samfunnsoppdraget sitt som er utarbeidelse av offisiell statistikk om Norge. Jeg vil studere hvordan SSB som behandlingsansvarlig[2] tolker og håndhever regelverket de er underlagt når

[2] Behandlingsansvarlig er definert i personvernforordningens artikkel 4 7): "*(...) en fysisk eller juridisk person, en offentlig myndighet, en institusjon eller ethvert annet organ som alene eller sammen med andre bestemmer formålet med behandlingen av personopplysninger og hvilke midler som skal benyttes (...)".*

det kommer til behandling av personopplysninger hvor formålet med behandlingen er til statistiske formål, med hovedfokus på anonymiseringen.

Jeg har begrenset oppgaven til kun å gjelde anonymisering i forbindelse med publisering av statistikk. SSB jobber også med forskning og analyse,[3] hvor de blant annet utleverer det de kaller mikrodata[4] til forskere. Jeg har ikke sett på arbeidet de gjør med anonymisering her.

Inntrykket i forkant og etter intervjuene var at SSB hovedsakelig forholder seg til de samme reglene når det kommer til anonymitet i statistikken. Derfor har jeg fokusert på statistikk generelt. Ingen av informantene jobbet operativt nok til å gå spesifikt inn på én type statistikk.[5] At all statistikk stiller de samme kravene til anonymitet innebærer i korte trekk at statistikken SSB utgir skal være anonym uavhengig av om oppgavegiver til statistikken er en fysisk eller juridisk person og uavhengig av type opplysning statistikken viser.

Under intervjuene med informantene fikk jeg mye informasjon rundt informasjonssikkerheten til SSB, dette kommer blant annet frem der jeg beskriver hvordan SSB behandler personopplysninger i forhold til personvernprinsippene. Noe av informasjonen jeg fikk skal ikke offentliggjøres og er i tillegg ute av scope for denne oppgaven. I kapittel 4.3 og 4.4.1 har jeg derfor utelatt noe informasjon.

[3] Se mer i kapittel 4.2.
[4] Mikrodata er metadata som benyttes til forskning.
[5] Se mer i kapittel 2.3.2.

2 Metode og case

2.1 Innledning

Denne oppgaven skrives ved det Juridiske fakultet, Avdeling for forvaltningsinformatikk. Studiet forvaltningsinformatikk er et tverrfaglig studium med juridisk, informatisk og samfunnsvitenskapelig vinkling. Forvaltningsinformatikk skal belyse de tverrfaglige problemstillingene eller aspektene på en måte som ikke de enkeltstående fagene vil dekke alene. Temaet for oppgaven og de tre forskningsspørsmålene tar for seg de tre elementene i faget forvaltningsinformatikk.

Jeg har valgt en kombinasjon av juridisk og samfunnsvitenskapelig metode i masteroppgaven. Den juridiske metoden benyttes for å finne frem og analysere gjeldende rett, mens den samfunnsvitenskapelige metoden kommer frem i innsamling av empiri og intervjuer med informanter. Kombinasjonen av de to metodene skal gi et grunnlag for analyse, drøftelse og konklusjon. Jeg går nærmere inn på de to ulike metodiske tilnærmingene i kapittel 2.2 og 2.3 under.

2.2 Juridisk metode

Peter Blume beskriver juridisk metode som en verktøykasse for juristene der formålet er å fastlegge hvilke rettsregler som kommer til anvendelse, og hvilke resultat dette medfører.[6] Rettskilder fungerer som juridiske argumenter hvor den juridiske metoden karakteriseres ved at alle relevante rettslige argumenter på en faglig korrekt måte er klarlagt og avveid mot hverandre for å gi svaret på et rettslig spørsmål.[7] For å finne frem til innhold i rettsreglene har jeg benyttet meg av den juridiske metoden rettskildelære.

Denne oppgaven er ikke ment å være sektorspesifikk men skal løfte generelle problemstillinger knyttet til behandling av personopplysninger og anonymisering gjennom å studere anonymiseringsprosessen til SSB. Hovedrettskildene i denne oppgaven er "Lov om

[6] Blume, 2006, s. 43.
[7] Blume, 2006, s. 51.

behandling av personopplysninger", herunder kalt personopplysningsloven og Lov om offisiell statistikk og Statistisk sentralbyrå, herunder kalt statistikkloven.

Den nye personopplysningsloven har gjort personvernforordningen til norsk rett ved inkorporasjon, slik dette er foreskrevet i EØS-avtalens artikkel 7 bokstav a. Den nye personvernforordningen erstattet da EUs personverndirektiv fra 1995. I Norge ble personopplysningsloven fra 2000 erstattet med en ny personopplysningslov 20. juli 2018. Det følger av personopplysningsloven § 1 at EUs personvernforordning, gjennom EØS-avtalen, blir en del av norsk rett. Personvernforordningen skal verne fysiske personer i forbindelse med behandling av personopplysninger og sørge for fri utveksling av slike opplysninger innad i EU. Siden personopplysningsloven henviser til personvernforordningen kommer jeg til å benytte meg av personvernforordningens bestemmelser i denne oppgaven med mindre særlige norske regler gjelder.

Mens personvernforordningen regulerer behandling av personopplysninger generelt, regulerer statistikkloven SSBs innsamling og bruk av opplysninger til statistiske formål. I tillegg regulerer statistikkloven SSBs virksomhet og organisering. Gjennom EØS-avtalen er Norge forpliktet til å utarbeide statistikk som er forankret i EUs statistikkprogram og gjennomført i norsk rett i forskrift til statistikkloven. SSB er dermed underlagt retningslinjer for europeisk statistikk. Andre rettskilder jeg har benyttet meg av er relevante forskrifter, fortalene til personvernforordningen og forarbeider.

Forordningen legger opp til en risikobasert tilnærming. I tillegg er forordningen teknologinøytral. Dette er utfordrende, og bidrar til at forordningen er vanskelig å tyde, og jeg får ikke mye håndfast ut av den. Jeg har også i stor grad benyttet meg av tilgjengelig informasjon som omfatter NOU-er, veiledere, informasjonsskriv og anbefalinger fra henholdsvis Datatilsynet og andre statlige aktører som Nasjonal Sikkerhetsmyndighet og Direktoratet for forvaltning og IKT. I tillegg til offentlig tilgjengelig informasjon har jeg fått innsyn i intern dokumentasjon fra SSB under intervjuene. Retningslinjer, anbefalinger og standarder utgitt av myndighetene på et gitt område kan sies å være en praktisk tolkning av rettsreglene.

Da jeg startet arbeidet med denne oppgaven var ikke personvernforordningen implementert i Norge eller resten av EU. Mye av bakgrunnsinformasjonen jeg leste meg opp på var dermed basert på den gamle personopplysningsloven. Eksempler på dette er retningslinjer (fra

engelsk: Opinon) utgitt av Artikkel 29-gruppen, som var EUs rådgivende organ på personvernspørsmål under personverndirektivet.[8] Bruken av eldre litteratur og rettskilder kunne ført til at jeg baserte meg på feilaktig regelverk og at retningslinjene ikke var gjeldende lengre. Da jeg fikk lest igjennom den norske oversettelsen av personvernforordningen, konkluderte jeg med at både personopplysningsbegrepet og personvernprinsippene[9]i personvernforordningen i hovedtrekk er tilsvarende gammel personopplysningslov. Datatilsynets veileder, *Anonymisering av personopplysninger* fra 2015, er i stor grad en oversettelse av Artikkel 29-gruppens Opinion 05/2014 on *Anonymisation Techniques*. Datatilsynet mener veilederen fortsatt er gjeldende og relevant på tross av at henvisningene viser til gammel personopplysningslov.[10]Artikkel 29-gruppen er nå erstattet med Personvernrådet.[11] Personvernrådet har adoptert en del av Artikkel 29-gruppens fortolkninger av personvernforordningen. Derimot er veilederne jeg har benyttet som kilde (opinion 05/2014 on *Anonymisation Techniques* og opinion 4/2007 on *The concept of personal data*) utarbeidet lenge før personvernforordningen, og dermed ikke oppdatert etter nytt regelverk.

I sluttfasen av arbeidet med masteroppgaven fremmet Finansdepartementet forslag til en ny statistikklov den 5.april 2019. Forslaget bygger på NOU fra året før (NOU 2018: 7 *Ny lov om offisiell statistikk og Statistisk sentralbyrå*). Jeg har basert oppgaven på dagens statistikklov, men nevner det nye forslaget enkelte steder i oppgaven der det er relevant.

Jeg oppfatter ikke personvernforordningen som helt entydig og, som jeg nevnte over har det vært utfordrende å tolke forordningen. Som et eksempel på dette trodde jeg selv at grensene for hva som var å regne som en personopplysning og hva som var en anonym opplysning var åpenbar. Likevel forble jeg ikke like sikker i min sak etter å ha fordypet meg i litteratur og rettskilder om temaet. Jeg synes for eksempel at det er utfordrende at personvernforordningen ikke er tydelig i hva som bør regnes som en personopplysning. I fortale 26 til personvernforordningen indikeres det at pseudonymiserte personopplysninger i enkelte tilfeller ikke er å regne som personopplysninger.[12]

[8] Direktiv 95/46/EF.
[9] I personopplysningsloven (LOV-2000-04-14-31) ble det som nå er omtalt som "prinsipper", omtalt som "grunnkrav til behandling av personopplysninger".
[10] Datatilsynet: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/informasjonssikkerhet/hvordan-anonymisere-personopplysninger/
[11] European Data Protection Board (EDPB)
[12] Mer om pseudonymisering og fortale 26 i kapittel 3.3.3

Forordningen legger dermed opp til at den behandlingsansvarlige selv må finne ut, ikke bare hva som er tilstrekkelig sikkerhet ved behandlingen, men også hva som er en personopplysning og hva som er en anonym opplysning. Mitt siste forskningsspørsmål er praktisk vinklet og ved bruk av intervju som metode[13] gjorde det mulig å finne ut av hvordan SSB som jobber med personopplysninger og anonymisering i det daglige tolker personvernforordningens bestemmelser.

2.3 Samfunnsvitenskapelig metode

Samfunnsvitenskapelig metode angir hvordan en skal fremskaffe og utvikle teorier i et bestemt fagområde.[14] Empiri er erfaringsbasert informasjon som er blitt samlet inn ved bruk av samfunnsvitenskapelige metoder. Jeg har benyttet meg av samfunnsvitenskapelig metode for å besvare forskningsspørsmål 2 og 3 som handler om hvilke teknikker som kan benyttes for best å ivareta personopplysningene til den registrerte og hva den behandlingsansvarlige kan gjøre for å sørge for at personopplysningene er sikret godt nok. Og til slutt, hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym.

Samfunnsvitenskapelig metode inneholder enten kvalitativ metode, som for eksempel intervjuer eller kvantitativ metode som for eksempel statistikk, ved bruk av spørreundersøkelser. Jeg har benyttet meg av kvalitativ metode der jeg har benyttet meg av intervjuer for å besvare spørsmål knyttet til temaet i denne oppgaven. Kjennetegnet på en kvalitativ metode er at intervjuer er knyttet tett opp til intervjuobjektet. I en casestudie, slik denne oppgaven er basert på, er det få enheter med mange variabler. Mitt undersøkelsesopplegg var semistrukturert og jeg benyttet meg av en intervjuguide under intervjuet. Intervjuguiden inneholdt en skisse over emner med forslag til spørsmål til hvert emne, dette for å sørge for at jeg fikk de svarene jeg trengte for å kunne besvare forskningsspørsmålene.[15]

Grønmo skiller på tre hovedtyper informasjonskilder i samfunnsvitenskapelige studier: aktør, respondenter og dokumenter. [16] I mitt tilfelle er aktøren jeg har studert SSB, respondentene mine er to informanter fra SSB, den ene er jurist og den andre statistiker. Dokumentene er

[13] Les mer om intervjuene og samfunnsvitenskapelig metode i kapittel 2.3.
[14] Grønmo, 2004 s. 28.
[15] Intervjuguiden er tilgjengelig som vedlegg i denne oppgaven.
[16] Grønmo, 2004, s. 120.

dokumentasjonen og informasjonen jeg har tilegnet meg gjennom denne masteroppgaven, inkludert intervjuene og rettskilder nevnt over. Som et kvalitativt undersøkelsesopplegg kan resultatet i denne oppgaven inneholde feilkilder og resultatene kan ikke generaliseres.

2.3.1 Informantene

Jeg har valgt ut informantene på to forskjellige nivåer i SSB. Den ene informanten har en overordnet rolle i SSB og den andre jobber mer spesifikt med statistikk. Det var min veileder som introduserte meg for muligheten for intervju med informant A da han tidligere har hatt tilknytning til Senter for rettsinformatikk.

Informant A har vært ansatt i SSB i 20 år som jurist og personvernombud. Han var det første personvernombud som ble utpekt i Norge i 2003. Informant A forteller at hans rolle i SSB innebærer å gi generell juridisk bistand, primært for ledelsen, men også innenfor hele SSB sitt virksomhetsområde. Informant A er også gruppeleder for juristene som innebærer et fagansvar, inkludert koordinering av oppgaver. Fagansvaret innebærer også det å skjære gjennom problemstillinger om nødvendig.[17]

Informant B er metodestatistiker i SSB med en doktorgrad i statistikk. Informant B bidrar med metoder for å sørge for at statistikk er anonym. Informant B forteller at han startet med disse metodene i folketellingen i 2001 deretter 2011. Metodene som ble brukt i folketellingen blir brukt i andre statistikker i dag. Som informant A har også Informant B lang fartstid i SSB.[18]

Begge informantene sitter i et internt utvalg i SSB som kalles Konfidensialitetsutvalget. Konfidensialitetsutvalget skal bidra med rådgivning på konfidensialitetsspørsmål. Informantene mine er dermed veldig bevisste på problemstillingen knyttet konfidensialitet generelt og til anonymitet i statistikk spesielt.

2.3.2 Gjennomføring av intervjuene

Formålet med intervjuene var å få en forståelse av hvordan SSB er organisert og hvordan de behandler personopplysninger som skal benyttes til statistikk internt, altså behandlingen av

[17] Intervju med informant A.
[18] Intervju med informant B.

personopplysninger før publisering av statistikk. I tillegg til de generelle kravene i personvernforordningen som SSB må forholde seg til, er det enkelte unntak når det kommer til behandling av personopplysninger for statistiske formål. Videre ville jeg få besvart forskningsspørsmål 3: *Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym?*

Artikkel 89 nr. 1 i personvernforordningen stiller noen krav til garantier som må følges for å være unntatt enkelte av bestemmelsene i forordningen. Jeg ønsker å undersøke hvordan SSB forholder seg til disse unntakene og garantiene etter artikkel 89 nr. 1. Som jeg nevnte i kapittel 2.2 trådte ny personopplysningslov i kraft den 20. juli 2018. Det var derfor naturlig å spørre hvorvidt SSB gjort store endringer i måten de behandler personopplysninger på, som følge av nytt regelverk.

For å forsikre at alle spørsmålene blir besvart, hadde jeg i forkant av intervjuet utarbeidet en intervjuguide. Intervjuet har vært semistrukturert ved at jeg har latt informanten snakke fritt uten å styre for eksempel detaljeringsgraden i svarene som blir gitt. For å finne ut av forskningsspørsmålene, forsøkte jeg å oppfordre informanten til å benytte egne ord med detaljert informasjon. Det var ønskelig at informanten reflekterte rundt spørsmålene som ble stilt, på denne måten unngikk jeg for mange føringer fra min side.

Jeg startet med å sende en forespørsel til informant A om han kunne stille til intervju etter å ha forklart temaet for masteroppgaven og bakgrunnen for at jeg ønsket intervju med ham. Vi hadde en telefonsamtale en uke etter hvor vi avklarte tid og sted for intervju. Informant A og jeg ble enige om at etter vårt møte ville informant A introdusere meg for én eller flere informanter som kunne svare på detaljerte spørsmål om teknikker og metoder for anonymisering.

I forkant av begge intervjuene ba jeg om samtykke til å ta lydopptak av intervjuet. Begge informantene samtykket til dette. Jeg valgte å transkribere begge intervjuene for å få med meg mest mulig informasjon. Videre laget jeg referat som jeg sendte ut til informantene for å få tilbakemelding og godkjennelse for at det som sto i referatet kunne brukes som kilde til denne masteroppgaven. I forbindelse med godkjenning av referatene ble det enkelte oppfølgingsspørsmål på hva som kunne benyttes som kilde, og hva som skulle være unntatt offentlighet.

Etter min mening gikk gjennomføringen av intervjuene veldig bra. Begge informantene var svært åpne og ærlige, både der de selv snakket fritt rundt et tema og ved konkrete spørsmål fra meg. Dette førte til en uformell tone hvor praten gikk lett. Den åpne tonen førte også til at, spesielt informant A, enkelte ganger sa ting som ikke nødvendigvis skal være offentlig tilgjengelig informasjon. Der informant A var litt for frittalende betrygget jeg ham med at han ville få tilsendt referat og hadde da mulighet til å fjerne informasjon som ikke var ment for offentligheten. Det er spesielt detaljer rundt informasjonssikkerheten til personopplysningene som ikke skal offentliggjøres. Selve anonymiseringen som prosess samt metodene og teknikkene de benytter er offentlig informasjon, med unntak av enkelte terskelverdier, omhandlet i kapittel 5.2.1. Jeg opplevede det som svært positivt at informantene var såpass åpne. Dette styrket troverdigheten deres og gjorde at jeg fikk bredere innsikt i SSBs tilnærming til behandling av personopplysninger. Hovedfokuset i denne oppgaven er ikke informasjonssikkerheten til SSB, men hvordan personopplysningene blir anonymisert. Dermed var det ikke problematisk at noe av informasjonen jeg fikk under intervjuene ikke er med som kilde.

I etterkant ser jeg at det hadde vært interessant å intervjue en tredje informant, en som jobber operativt med utarbeidelse av en spesifikk personstatistikk. Dette for å se helt konkret hva som skjer fra de lagrede personopplysningene med fødselsnummer som unik identifikator, blir til tall i en tabell. Jeg spurte informant B om de hadde dokumentasjon eller en form for prosesskart tilgjengelig for å forstå hele prosessen, men dessverre hadde de ikke det. Av praktiske erfaringer jeg har gjort, var det svært tidskrevende å transkribere intervjuene. Til sammen varte intervjuene over tre timer noe som tilsvarte 18 sider tekst. Referatene ble på ca. fem sider per intervju. Det ble sagt mye på intervjuene som ikke er like relevant for oppgaven og unødvendig å ha med i referatet, og jeg kunne spart tid på å kun skrive referat.

2.4 Oversikt over den videre fremstillingen

I kapittel tre går jeg inn på den juridiske forståelsen av hva personopplysninger er. Det er en forutsetning å ha kjennskap til og en forståelse for hva en personopplysning er, for å kunne vurdere hva som er en anonym opplysning. Videre i kapittelet introduserer jeg personvernprinsippene. Personvernprinsippene kan sees på som grunnleggende krav som må foreligge for å kunne behandle personopplysninger. Til slutt i kapittel tre gjennomgår jeg flere beskyttelsesteknikker som kan benyttes for å ivareta personopplysningssikkerheten og

personvernprinsippene. Jeg illustrerer en linje med to ytterpunkter av personopplysninger. Der det på den ene siden er snakk om entydig identifiserende personopplysning og på den andre siden er snakk om en anonym opplysning. I kapittel tre svarer jeg på forskningsspørsmål 1 og 2: *Hvilke rettsregler gjelder for identifisering og for beskyttelse av identiteter knyttet til personopplysninger* og *hvilke teknikker kan benyttes for å beskytte de identifiserende elementene?*

I kapittel fire introduseres Statistisk sentralbyrå. Her beskriver jeg SSBs samfunnsoppdrag og organisering. Videre går jeg inn på behandlingsgrunnlaget til behandling av personopplysninger som SSB har. I slutten av kapittelet ser jeg på hvordan SSB behandler personopplysninger internt med spesiell vekt på hvordan de samler inn og lagrer personopplysningene de skal bruke til statistikken sin. Dette for å se om behandlingen av personopplysninger underbygger personvernprinsippene omtalt i kapittel tre.

I kapittel fem går jeg inn på krav til konfidensialitet i statistikken som SSB er underlagt. I dette kapittelet kommer SSB sine interne vurderinger rundt anonymitet i statistikken frem. I visse unntakstilfeller kan SSB publisere statistikk som ikke er helt anonym, i kapittel fem beskrives prosessen der SSB vurderer om denne statistikken skal publiseres eller ikke.

I kapittel seks går jeg mer i dybden og ser på metoder og teknikker SSB bruker for å anonymisere statistikken. I kapittel fem og seks besvares forskningsspørsmål nummer 3: *Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at opplysningene er og forblir anonyme?*

I kapittel syv kommer jeg med en oppsummering og avsluttende kommentarer på masteroppgaven. Helt til slutt har jeg med noen tanker om videre arbeid som muligens vil være av interesse for andre som interessenter seg for problemstillinger knyttet til personvern og anonymisering.

3 Personopplysninger og identifisering

3.1 Grunnleggende om personvernbegrepet

For å forstå personvernregelverket må man vite hva en personopplysning er. Begrepet personopplysning står definert i personvernforordningen som:

> «personopplysninger» enhver opplysning om en identifisert eller identifiserbar fysisk person (…).[19]

En forståelse av personopplysningsbegrepet gjør at man kan skille mellom hva som faller innenfor eller utenfor personvernregelverket. Personvernforordningens saklige virkeområde står definert i artikkel 2 nr. 1: "Denne forordning får anvendelse på helt eller delvis automatisert behandling av personopplysninger og på ikke-automatisert behandling av personopplysninger som inngår i eller skal inngå i et register." I artikkel 2 nr. 2, om lovens saklige virkeområde, nevnes enkelte kontekster der forordningen ikke får anvendelse. Eksempelvis gjelder ikke personvernforordningen som ledd i rent personlige eller familiemessige aktiviteter om behandlingen utføres av en fysisk person.

Personopplysningsloven med personvernforordningen regulerer behandlingen av personopplysninger og personvernforordningen skal sikre vern av fysiske personers grunnleggende rettigheter og friheter, særlig deres rett til vern av personopplysninger.[20] Forordningens formål fremgår av artikkel 1, og forordningen gjelder opplysninger om fysiske personer. Dette innebærer at juridiske personer faller utenom lovens saklige område. I tilfeller der opplysninger om juridiske personer også sier noe om en fysisk person kan disse opplysningene være underlagt personvernforordningen. For eksempel dersom en psykolog driver et enkeltmannsforetak og kontoret ligger i privatboligen til eier kan denne adressen regnes som en personopplysning. Enkelt forklart er altså en personopplysning en opplysning som *kan* knyttes til en bestemt person.

[19] Personvernforordningen artikkel 4 nr. 1.
[20] Personvernforordningen artikkel 1 nr. 1 og 2.

Med `behandling av personopplysninger` menes alt som gjøres med disse personopplysningene, enten det er automatisert eller ikke. Personvernforordningen nevner mange eksempler på behandling av personopplysninger i artikkel 4 nr. 2, noen av disse er: "(…) innsamling, registrering, organisering, strukturering, lagring, tilpasning eller endring, gjenfinning, konsultering, bruk, utlevering ved overføring (…)".[21]

Innenfor lovens saklige virkeområde nevnt ovenfor har jeg vanskeligheter med å se for meg at en kan foreta seg noe med en personopplysning som ikke er å anse som `en behandling`. Videre definerer personvernforordningen den personen som personopplysningene er om for `den registrerte`.[22]

For å finne ut om en opplysning er en personopplysning i henhold til personvernregelverket eller om det er en anonym opplysning kan personopplysningsbegrepet brytes ned til tre deler som kan vurderes nærmere.[23] Kriteriene for at en personopplysning ikke er en anonym opplysning er som følgende:

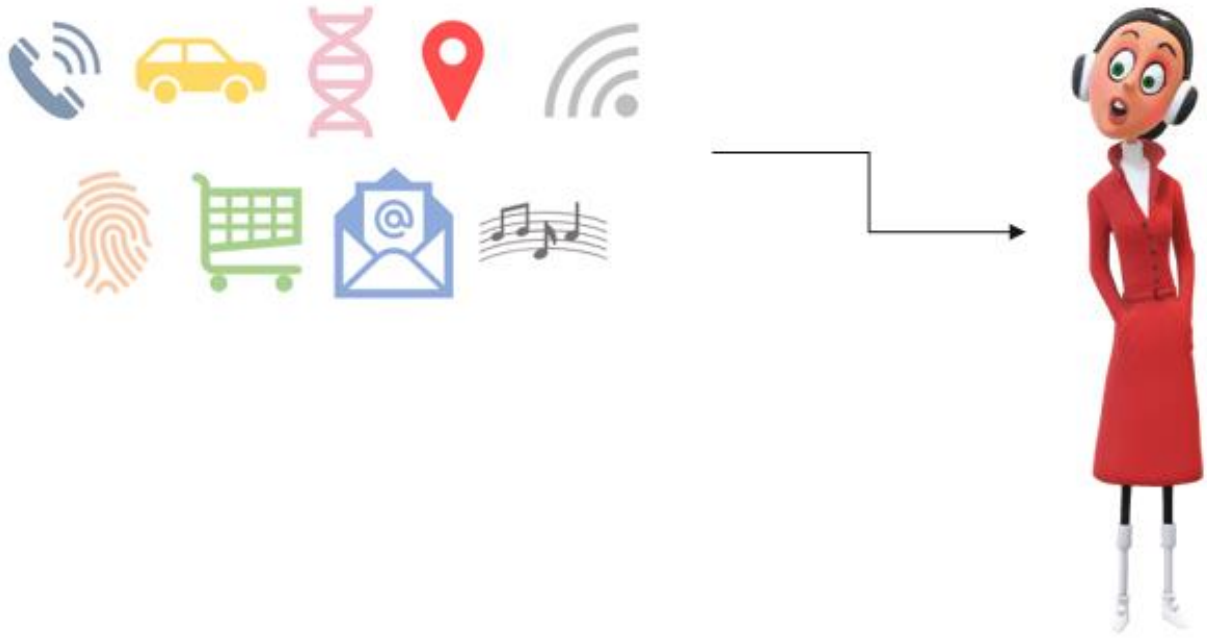
1) Enhver form for informasjon

2) Tilknytningselementet

3) Identifiserbar eller identifisert enkeltperson

Figuren under skal illustrere diverse opplysninger som kan være personopplysninger.

[21] Personvernforordningen artikkel 4 nr. 2.
[22] Personvernforordningen artikkel 4 nr. 1.
[23] Artikkel 29-gruppen: Opinion 05/2014 on *Anonymisation Techniques*.



Figur 1 Illustrasjon av potensielle personopplysninger

Ikonene i figuren skal illustrere, fra venstre: telefonnummer, bilskilt, DNA, lokasjon, signal fra en nettverkskomponent som for eksempel IP-adresse eller MAC-adresse[24], fingeravtrykk, varer en person har handlet og koblet opp mot identitet[25], e-postadresse og sangstemme i form av lydbølger.

Både objektive og subjektive opplysninger om en person vil være en personopplysning. Eksempler på objektive opplysninger kan være: telefonnummer, en persons høyde eller arbeidsgiver. Med subjektive opplysninger menes noens vurderinger eller tolkninger av en person; det trenger ikke være realitet eller fakta knyttet til opplysningen. Påstanden: "sjefen min er uselvstendig", er et eksempel på en subjektiv opplysning, da `uselvstendig` er en subjektiv oppfatning av informasjon. Det har ingen betydning hvilket format opplysningen kommer i. Det kan være tall, skrift, tegning eller for eksempel i form av kode.

Når informasjonen knyttes opp mot en bestemt person kalles dette en identifisering. Identifisering er en prosess der identifikasjonsmidler blir knyttet opp mot en identitet. I eksempelet over, der informasjonen er et telefonnummer, vil dette nummeret være knyttet til en person, enten om knytningen forekommer internt hos en telefonoperatør eller om dette er

[24] "En MAC-adresse er et unikt identifikasjonsnummer som kan identifisere din mobiltelefon. Det er tildelt av produsenten. En MAC-adresse defineres som en personopplysning når den samles inn gjennom WiFi-sporing". Hentet fra Datatilsynets veileder: *Sporing i det offentlige rom* (2016) s.4.
[25] For eksempel gjennom Trumf-medlemskap eller lignende tjenester.

offentlig tilgjengelig informasjon. Ved å identifisere noen fastslår en hvem personen er. Det kan forekomme en viss usikkerhet i identifiseringen. I telefonnummer-eksempelet kan telefonnummeret tilhøre en person med navn Ola Nordmann, denne informasjonen er offentlig tilgjengelig, men det finnes flere Ola Nordmann i Norge. Ved hjelp av den offentlige tilgjengelige informasjonen om telefonnummer og navn kan en ikke, uten annen informasjon, være helt sikker på at telefonnummeret tilhører en bestemt person. Telefonselskapet derimot har behov for å vite akkurat hvilken person som har dette nummeret for å kunne opprette kundeforhold, kunne fakturere osv. Telefonselskapet sitter derfor på mer informasjon om Ola Nordmann og har et behov for å entydig identifisere Ola Nordmann. Mer om entydig identifisering i kapittel 3.3.1.

I tilfeller der det er mange som har samme navn er det behov for flere opplysninger for å kunne identifisere en person. Men ut ifra en viss kontekst vil også et vanlig fornavn kunne være direkte identifiserende. En kan argumentere for at et av de mest brukte navnene ikke kan kalles en personopplysning om navnet er eneste opplysning.

Personopplysningsvernet skiller på type opplysning som behandles og behandling av *særlige kategorier* av personopplysninger er som hovedregel forbudt.[26] På tross av at denne reglen har mange unntak er det tydelig at enkelte typer personopplysninger er av en mer sensitiv karakter og har behov for et sterkere vern enn andre personopplysninger, enten om dette er i form av lovhjemler eller beskyttelsesteknikker. Når det kommer til sikkerhet ved behandlingen av personopplysninger vil det være særlig viktig at den behandlingsansvarlige tar hensyn til hvilke typer opplysning som behandles.

[26] Personvernforordningen artikkel 9 nr. 1.

3.2 Prinsipper for behandling av personopplysninger

Personvernforordningens artikkel 5 definerer syv personvernprinsipper for behandling av personopplysninger. Prinsippene kan man se på som grunnleggende krav til hva som må til for at det er lov å behandle personopplysninger. Under vil jeg gi en kort beskrivelse av hva prinsippene innebærer og hvordan prinsippene er gjeldende når formålet for behandlingen er statistikk.

Figuren under skal illustrerer personvernprinsippene etter personvernforordningens artikkel 5.

Figur 2 Personvernprinsippene.

Det første prinsippet, "**Lovlig, rettferdig og åpenhet**", går ut på at den som behandler personopplysningene må ha et behandlingsgrunnlag (lovlig) .[27] Et grunnlag for å behandle personopplysninger kan for eksempel være lovhjemmel eller gjennom samtykke. SSBs behandling av personopplysninger er hjemlet i statistikkloven.[28] I tillegg innebærer prinsippet at behandlingen av personopplysninger må skje rettferdig. Det betyr at behandlingen skjer gjennom den registrertes interesser og rimelige forventninger. Prinsippet om åpenhet krever at

[27] Personvernforordningens artikkel 6 nr.1 beskriver behandlingens lovlighet.
[28] Mer om rettslig grunnlag for behandling av personopplysninger for SSB i kapittel 4.3.

all informasjon og kommunikasjon i forbindelse med behandling av nevnte personopplysninger er lett tilgjengelig og lettfattelig.[29]

Prinsippet om "**formålsbegrensning**" går ut på at personopplysningene som samles inn skal samles inn for spesifikke, uttrykkelige, angitte og legitime formål. Formålet skal forklares på en slik måte at alle som er berørt har en forståelse av hva personopplysningene skal brukes til. En kan tenke seg at formålet: "Bedre helsevesen" ikke er spesifikt nok, og med et slikt upresist formål er det ikke klart hvilke personopplysninger som må behandles for å kunne oppnå formålet. Derimot er formålet: "Administrere lønnen til ansatte" et mer spesifikt formål hvor det er mulig for den enkelte å forestille seg hvilke opplysninger som må behandles for utbetaling av lønn. Om formålsbegrensning står det i personvernforordningens artikkel 5 bokstav b at:

> "(..) viderebehandling for arkivformål i allmennhetens interesse, for formål knyttet til vitenskapelig eller historisk forskning eller for statistiske formål skal, i samsvar med artikkel 89 nr. 1, ikke anses som uforenlig med de opprinnelige formålene («formålsbegrensning»)".

Dette innebærer at statistikk ikke skal ansees som uforenelig med de opprinnelige formålene. Med dette følger det noen garantier etter artikkel 89 nr. 1. Den behandlingsansvarlige må da "sikre at det er innført tekniske og organisatoriske tiltak for særlig å sikre at prinsippet om dataminimering overholdes".[30] Pseudonymisering som omtales i kapittel 3.3.3 er eksempel på en slik garanti.

Videre kommer prinsippet om "**dataminimering**", som innebærer at de opplysningene som samles inn skal være adekvate og relevante for formålene de behandles for.[31] Hvis ikke opplysningene som samles inn, eller behandles, er relevante for å oppnå formålet, skal de heller ikke behandles. Som nevnt over skal ikke statistikk skal være uforenelig med de opprinnelige formålene, men da må den behandlingsansvarlige blant annet sørge for at prinsippet om dataminimering overholdes.

Prinsippet om "**riktighet**" innebærer at opplysningene skal være korrekte og oppdaterte. Der avgjørelser kan bli tatt på bakgrunn av opplysningene er det svært viktig at opplysningene er

[29] Personvernforordningen fortale 39.
[30] Jf. personvernforordningen artikkel 89 nr. 1.
[31] Personvernforordningen artikkel 5 nr. 1 bokstav c.

riktige. Prinsippet "riktighet" underbygger flere av rettighetene til den registrerte etter personvernforordningens tredje kapittel, blant annet rett til retting.[32] For behandling av personopplysninger for statistikkformål finnes det enkelte unntak fra personvernforordningen når det kommer til den registrertes rettigheter. [33]

"**Lagringsbegrensning**" innebærer at opplysningene ikke skal lagres lengre enn nødvendig for å oppnå formålet med behandlingen. For statistiske formål åpner prinsippet om lagringsbegrensning opp for at personopplysningene kan lagres i lengre perioder såfremt garantiene etter artikkel 89 nr. 1 er på plass og forutsatt at det "gjennomføres egnede tekniske og organisatoriske tiltak som kreves i henhold til denne forordning for å sikre de registrertes rettigheter og friheter".[34]

Prinsippet om "**integritet og fortrolighet**" innebærer blant annet at opplysningene skal behandles på en måte som sikrer tilstrekkelig sikkerhet for opplysningene. Den behandlingsansvarlige har ansvar for å sørge for at personopplysningene er ivaretatt i henhold til integritet, tilgjengelighet og konfidensialitet. Dette prinsippet gjenspeiler krav til sikkerhet til behandlingen av personopplysningene. Et eksempel på et teknisk sikkerhetstiltak for å sikre konfidensialitet er at uautoriserte forsøk på pålogging i et system blir registrerte og logget. At den behandlingsansvarlige har en prosess for å håndtere uønskede hendelser kan også være et organisatorisk sikkerhetstiltak for å sikre prinsippet om integritet og fortrolighet. Sikkerhetstiltak som pseudonymisering og kryptering som jeg omtaler i kapittel 3.3.3 og 3.3.4 vil kunne bidra til at prinsippet om integritet og fortrolighet blir ivaretatt.

Det siste personvernprinsippet er ifølge personvernforordningen prinsippet om "**ansvarlighet**". Den behandlingsansvarlige skal kunne påvise at personvernprinsippene etterleves. I Datatilsynets veileder *Grunnleggende personvernprinsipper*, beskrives prinsippet ansvarlighet ved at en behandlingsansvarlig må opptre proaktivt og etablere alle nødvendige organisatoriske og tekniske tiltak for å sikre at regelverket etterleves til enhver tid.[35]

[32] Personvernforordningen artikkel 16.
[33] Personopplysningsloven § 17.
[34] Personvernforordningen artikkel 5 nr. 1 bokstav e.
[35] Datatilsynets veileder: *Grunnleggende personvernprinsipper*.

3.3 Oversikt over ulike grader av identifisering

Som nevnt i kapittel 3.1 må tilknytningselementet være til stedet for at vi har med en personopplysning å gjøre. Tilknytningselementet vil kunne være åpenbart, som for eksempel i et kunderegister eller informasjon på en lønnsslipp. Men tilknytningselementet kan også være vanskelig å tyde, for eksempel når det kommer til opplysninger om en gjenstand. Hvis en ser for seg informasjon om en bil: Ved første øyekast er dette nettopp opplysninger om en bil, men med flere variabler og type informasjon om bilen vil denne informasjonen kunne knyttes opp til en enkeltperson. Opplysning om farge og for eksempel antall kilometer kjørt vil nok neppe kunne knyttes til en enkeltperson. Sammenstilles denne informasjonen med registreringsnummeret på bilen, eller lokasjonsopplysninger på hvor bilen befinner seg, har man derimot med en personopplysning å gjøre. En personopplysning kan i visse tilfeller også være personopplysninger om andre enn den enkelte, for eksempel kan biometriske opplysninger si noe om genetikken til en person. Personopplysningene vil derfor kunne si noe om andre familiemedlemmer. Biometriske opplysninger om en person vil derfor indirekte være opplysninger om denne personens bror. Det vil si at opplysninger som gjennom flere ledd kan kobles opp mot en enkelt person er å regne som en personopplysning.

At en person er identifiserbar betyr at denne personen kan skilles ut i fra en gruppe med personer. Personen trenger dermed ikke å være identifisert, men muligheten må være der. Personen må være identifiserbar. For at et individ skal kunne identifiseres er det behov for en eller flere opplysninger om denne personen. Thomas Olsen beskriver at identitet i dagligtalen gjerne er synonymt med den enkeltes alminnelige brukte navn.[36] Opplysningene som gjør at en kan identifisere en person må være opplysninger som skiller personen fra andre personer, disse opplysningene blir kalt identifikasjonsmidler. En identifisering er avhengig av en eller flere identifikasjonsmidler. Identifisering er derfor en prosess der identifikasjonsmidler blir knyttet opp mot en identitet.

Et eksempel på en opplysning der det er usikkerhet rundt tilknytningselementet er en IP-adresse. En IP-adresse (identifikator til en gjenstand tilknyttet internett[37]) vil være en personopplysning der en PC blir benyttet i en husstand, mens en IP-adresse til en PC på et bibliotek neppe kan betraktes som en personopplysning alene. Dette bekrefter også en EU-

[36] Olsen, 2015, s. 109.
[37] Schartum og Bygrave, 2016, s. 139.

dom som fastslår at dynamiske IP-adresser er å regne med som en personopplysning.[38] Det er altså tilstrekkelig at identifikasjonen av individet kan tenkes å finne sted en gang i fremtiden for at personopplysningsloven kommer til anvendelse. I tillegg kan den mulige identifikasjonen utføres av noen andre enn den behandlingsansvarlige selv. Derfor kan en behandling av personopplysninger som den behandlingsansvarlige har vurdert til å være anonyme opplysninger i visse tilfeller innebære en ulovlig behandling av personopplysninger.

Det at en tilsynelatende anonym opplysning allikevel vil kunne identifiseres, eller reidentifiseres, gjør det utfordrende å garantere at en opplysning er, og ikke minst at den vil forbli, anonym. Ved bruk av stordata hvor store mengder data fra forskjellige kilder sammenstilles, kan man forestille seg at en reidentifisering kan skje en gang i fremtiden. Dessverre finnes det ikke en glasskule behandlingsansvarlige kan se i som vil fortelle dem om de anonymiserte dataene de har tilgjengeliggjort senere vil være gjenstand for en reidentifisering. Det en behandlingsansvarlig kan gjøre er grundige vurderinger i forkant av behandlingen for å avdekke potensielle risikoer knyttet til behandlingen. Ved å forhåndsdefinere uønskede hendelser og vurdere sannsynlighet og konsekvens for at en slik hendelse kan inntreffe, kan den behandlingsansvarlige på en noenlunde vitenskapelig måte komme frem til et resultat med en risiko som er akseptabelt.[39] I tilfeller der behandlingen kan medføre en høy risiko for den registrerte er den behandlingsansvarlige pliktet til å utføre en vurdering av personvernkonsekvensene etter personvernforordningen artikkel 35.

Videre i denne delen av kapittelet skal jeg beskrive hvordan en behandlingsansvarlig kan behandle personopplysninger ved bruk av entydige identifikasjonsmidler til å benytte forskjellige teknikker for å beskytte personopplysningene. Figuren under skal illustrere en linje som går fra å benytte seg av entydige identifiserte personopplysninger helt til venstre, igjennom en del teknikker for å beskytte personopplysningene, over til høyre side der det ikke lenger er snakk om en personopplysning da personopplysningen er blitt anonymisert.

[38] Judgment in Case C-582/14.
Patrick Breyer v Bundesrepublik Deutschland: https://curia.europa.eu/jcms/upload/docs/application/pdf/2016-10/cp160112en.pdf
[39] Hentet fra nettsidene til DIFI som gjengir ISO/IEC 27001- Information security management. Risikoidentifisering, risikoanalyse og risikoevaluering.
https://internkontroll-infosikkerhet.difi.no/risikostyring/risikovurdering

Figur 3 Ulike grader av identifisering

Fra venstre er `Entydig identifisert` illustrert med et ikon for en DNA-streng. `Avidentifisert` illustreres med to personer da opplysningene ikke direkte kan knyttes opp mot en enkelt person, eller det er en viss usikkerhet om opplysningene tilhører en enkelt person. `Pseudonymisert` er illustrert med en liste over pseudonymer. `Kryptert` er illustrert med en hengelås og til slutt er `Anonymisert` illustrert med flere personer som skal indikere at opplysningen ikke kan knyttes opp mot en enkelt person. Den loddrette linjen illustrerer et skille der opplysningen er innenfor eller utenfor personopplysningsregelverket.

3.3.1 Entydig identifiserende opplysninger

Entydig identifiserende opplysninger kan deles opp i begrepene *entydig* og *identifiserende*. I kapittel 3.3 omtalte jeg ulike grader av identifisering. Identifisering er som oftest betegnelse på prosessen å fastslå en persons identitet[40] og et av kravene til at en opplysning er en personopplysning går på at muligheten for en identifisering er tilstedte. Entydig innebærer en `felles forståelse`, at noe er `utenfor enhver tvil` eller `kan ikke misforståes`. Entydig er altså noe som ikke kan tolkes på en annen måte. En entydig identifiserende opplysning er dermed en personopplysning som ved sikkerhet bekrefter en persons identitet.

I felles datakatalog blir `entydig identifisering av person` av Skatteetaten definert som: "identifisering med resultat at man er helt sikker på identiteten til en person."[41] Entydig identifikasjonsmidler er derfor noe annet enn navn, adresse, hårfarge eller kjønn.

[40] Schartum, *Utredning om fødselsnummer, fingeravtrykk og annen bruk av biometri i forbindelse med lov om behandling av personopplysninger § 12*, 2008, s. 8.
[41] Brønnøysundregistrene: https://fellesdatakatalog.brreg.no/.

Fødselsnummer som tildeles alle som er folkeregistret i Norge[42]er et eksempel på en entydig identifiserende opplysning.

Personopplysningsloven § 12 setter begrensninger i bruken av entydige identifikasjonsmidler. Det er ikke tillatt bruk av entydige identifikasjonsmidler uten at det er saklig behov for sikker identifisering. Entydig identifikasjonsmidler betyr at en eller flere personopplysninger er knyttet opp mot et entydig identifikasjonsmiddel slik at det ikke er tvil om at opplysningen tilhører den personen den er ment å kobles til. Det er kun entydige identifikasjonsmidler som egner seg til å identifisere en person og dermed gjør det mulig å skille en person fra en annen person. Dette innebærer at det i enkelte tilfeller vil være lov å behandle entydige identifikasjonsmidler nettopp for å kunne identifisere noen med sikkerhet, men personopplysningsloven § 12 skal sørge for at fødselsnummer og andre entydige identifikasjonsmidler ikke misbrukes eller brukes mer enn nødvendig.

Biometri gjennom biometriske kjennetegn betegnes også som entydig identifikasjonsmidler. Eksempel på dette kan være fingeravtrykk, iris eller DNA. Andre biometriske kjennetegn kan være en persons ganglag, hvordan personen benytter et tastatur eller for eksempel ansiktsformen til en person. Personvernforordningen har kategorisert biometriske opplysninger med formål å entydig identifisere noen som "særlige kategorier av personopplysninger."[43] Hovedregelen er at det er forbudt å behandle slike opplysninger. Selv om det finnes en rekke unntak til hovedregelen legger forordningen her til at visse former for entydige identifikasjonsmidler har behov for ekstra vern. Personvernrelaterte utfordringer knyttet til økt bruk av biometri blir trukket frem i NOU 2015:13 *Digital sårbarhet – sikkert samfunn*. Teknologien for ansiktsgjenkjenning beskrives som så god at det "snart vil være mulig å følge med på alle som beveger seg i et kameraovervåket område."[44] Bruk av fingeravtrykk som erstatning for passord løftes også som en økende trend.[45]

Forordningen nevner ikke konkrete tilfeller der det vil være saklig behov for sikker identifisering. Med en sikker identifisering ved bruk av entydige identifikasjonsmidler vil den behandlingsansvarlige i visse tilfeller sørge for tilstrekkelig sikkerhet ved behandlingen, spesielt med tanke på integriteten og konfidensialiteten til opplysningene. Det er dermed en

[42] Skatteetaten: www.skatteetaten.no/person/folkeregister/fodsel-og-navnevalg/barn-fodt-i-norge/fodselsnummer/
[43] Jf. personvernforordningen artikkel 9 nr. 1.
[44] NOU 2013:13 s. 50.
[45] NOU 2013:13 s. 50.

balansegang mellom å sikre seg identiteten til en person opp mot å ikke misbruke entydige identifikasjonsmidler da konsekvensene for brudd på personopplysningssikkerheten vil være alvorligere om for eksempel entydige identifikasjonsmidler kommer på avveie og blir misbrukt. Bruk av fødselsnummer som entydig identifiserende opplysning er nødvendig i flere sammenhenger som for eksempel med å rapportere til skattemyndighetene. Dette innebærer at det i flere tilfeller vil være lov å behandle entydige identifikasjonsmidler nettopp for å kunne identifisere noen med sikkerhet, men personopplysningsloven § 12 skal sørge for at fødselsnummer og andre entydige identifikasjonsmidler ikke misbrukes eller brukes mer enn nødvendig.

Det er viktig å få frem at bruk av entydig identifiserende personopplysninger ikke trenger være en ulempe for personvernet, tvert imot vil det å benytte entydige identifikasjonsmidler i forbindelse med en behandling av personopplysninger kunne sikre personvernprinsippet om integritet og fortrolighet og dermed ivareta enkelte av den registrertes rettigheter etter personvernforordningens kapittel 3. På grunn av den teknologiske utviklingen foreligger det en bekymring i at stor spredning av unike identifikasjonsmidler, øker sannsynligheten for misbruk som kan føre til store konsekvenser for den enkelte, som for eksempel ID-tyveri.[46] Ved økt bruk av biometri som identifikasjonsmiddel nevnt ovenfor kan en se for seg at konsekvensene for den enkelte vil være høyere om denne personens biometriske opplysninger kommer på avveie og blir misbrukt. Satt på spissen vil det være enklere å bytte ut brukernavn og passord enn å bytte fingeravtrykk.

Mens entydige identifikasjonsmidler, som for eksempel fødselsnumre, har noen begrensninger i bruk, finnes det mange andre opplysninger som kan identifisere en person. Entydige identifikasjonsmidler kan også sees på i en gitt kontekst. For eksempel vil fødselsnummer være entydig identifiserende i konteksten "innbyggere i Norge". Innad i en virksomhet kan for eksempel brukernavnet som benyttes for å koble opp mot en enhet[47] være entydig identifiserende.

[46] "En identitetstyv kan lykkes med mindre svindelforsøk om han eller hun har tilgang til noens personopplysninger slik som navn, adresse, fødselsnummer og e-post" Hentet fra Datatilsynet: https://www.datatilsynet.no/personvern-pa-ulike-omrader/internett-og-apper/id-tyveri/identitetstyveri--hva-trenger-id-tyven-og-hvordan-beskytter-du-deg---/

[47] For eksempel brukernavnet i Active Directory, som er Microsoft sin katalogtjeneste for å blant annet kunne håndtere brukere og rettigheter.

3.3.2 Avidentifisering

Avidentifisering innebærer at det å identifisere et individ er gjort vanskeligere ved at de identifiserende opplysningene er fjernet. Det vil fortsatt være mulig å koble opplysningene opp mot en enkeltperson, dermed er en avidentifisert personopplysning fortsatt en personopplysning. Avidentifiserte personopplysninger er derfor en form for indirekte identifiserbare personopplysninger. Avidentifiserte personopplysninger skiller seg fra pseudonymisering som omtales i kapittel 3.3.3 ved at de identifiserende elementene er fjernet, mens i pseudonymisering er de identifiserende elementene endret eller byttet ut med en annen identifikator.

Personvernforordningen nevner ikke avidentifisering direkte, men Datatilsynet definerer begrepet avidentifisering som en personopplysning der navn, personnummer eller andre personentydige kjennetegn er fjernet.[48] Begrepet avidentifisering blir benyttet i flere forskrifter innunder helselovgivningen.[49] I forskrift om innsamling og behandling av helseopplysninger i Norsk pasientregister (Norsk pasientregisterforskriften) blir avidentifiserte opplysninger definert som:

> "(…) helseopplysninger der navn, fødselsnummer og andre personentydige kjennetegn er fjernet, slik at opplysningene ikke lenger kan knyttes til en enkeltperson, og hvor identitet bare kan tilbakeføres ved sammenstilling med de samme opplysninger som tidligere ble fjernet."[50]

Det å avidentifisere et sett med personopplysninger vil derfor være et tiltak som begrenser risikoen knyttet til behandlingen, da opplysninger på avveie vil ha lavere konsekvenser for den registrerte, i tillegg til at avidentifisering er med på å underbygge prinsippet om dataminimering i henhold til personvernforordningen artikkel 5 nr. 1 bokstav c.

I henhold til artikkel 11 i personvernforordningen skal ikke den behandlingsansvarlige, dersom formålene med behandling av personopplysninger ikke krever det, ha plikt til å bevare, innhente eller behandle ytterligere opplysninger for å identifisere den registrerte utelukkende med det formål å oppfylle kravene i forordningen. Artikkel 11 underbygger

[48] Hentet fra Datatilsynet sin ordbok: https://www.datatilsynet.no/regelverk-og-verktoy/verktoy/ordbok-a-til-a/
[49] Blant annet i Norsk pasientregisterforskrift, kreftregisterforskriften og forskrift om overføring av biomateriale til utlandet.
[50] Norsk pasientregisterforskrift § 1-5 bokstav a.

kravet til dataminimering nevnt i kapittel 3.2 og oppfordrer dermed den behandlingsansvarlige til å benytte avidentifisering dersom det er mulig. Dette på tross av at en avidentifisering kan føre til at det blir vanskelig å ivareta kravene til enkelte av den registrertes rettigheter i artikkel 15-20.[51]

3.3.3 Pseudonymisering

Ordet pseudonym stammer fra gresk og betyr falskt navn. Identifikatoren `navn` er byttet ut med et fiktivt navn. Det finnes mange eksempler på bruk av pseudonymer, først og fremst kanskje blant forfattere. Henrik Ibsen skrev i perioder under pseudonymet Brynjolf Bjarme mens Charles Dickens brukte pseudonymet Boz. Det finnes utallige eksempler på personer som har benyttet seg av pseudonymer, også den fiktive karakteren Bruce Wayne er nok bedre kjent under pseudonymet Batman. Det å benytte seg av et falskt navn er også aktuelt for dem som ønsker å for eksempel ytre seg uten å røpe sin identitet. Falske profiler eller profiler under et fiktivt navn er ikke unormalt i sosiale medier.

Et navn eller et annet identifiserende element som er byttet ut med et pseudonym vil fortsatt være en personopplysning da det finnes en mulighet å koble pseudonymet til en fysisk person på en eller annen måte. I eksempelet over er det rimelig å tenke at forleggeren til en forfatter som bruker pseudonym er kjent med forfatterens identitet. Pseudonymisering blir derfor benyttet som et risikoreduserende tiltak for de registrerte og en mulig løsning på hvordan behandlingsansvarlig og databehandler skal kunne oppfylle sine plikter etter personopplysningsloven. Artikkel 29-gruppen peker på skillet mellom to-veis og én-veis pseudonymisering.[52] To-veis pseudonymisering innebærer at man skal kunne reidentifisere ved å koble pseudonymet med opprinnelig identifikator. En-veis pseudonymisering derimot innebærer at man ikke kan reidentifisere personene som opplysningene stammer fra.[53] En kan dermed argumentere for at en-veis pseudonymisering i visse tilfeller ikke lenger er en personopplysning.

Pseudonymisering er definert i personvernforordningen:

[51] Personvernforordningen artikkel 11 nr. 2.
[52] Artikkel 29-gruppen: Opinion 4/2007 *on the concept of personal data*, 2007, s 19.
[53] Olsen, 2015, s. 163.

" «pseudonymisering» behandling av personopplysninger på en slik måte at personopplysningene ikke lenger kan knyttes til en bestemt registrert uten bruk av tilleggsopplysninger, forutsatt at nevnte tilleggsopplysninger lagres atskilt og omfattes av tekniske og organisatoriske tiltak som sikrer at personopplysningene ikke kan knyttes til en identifisert eller identifiserbar fysisk person."[54]

Pseudonymisering er ikke en form for anonymiseringsteknikk, men et sikkerhetstiltak der enkelte identifiserbare parameter blir erstattet med pseudonymer. Det finnes dermed en mulighet for at individet kan indirekte bli identifisert ved bruk av pseudonymer. Dette innebærer at selv om personopplysningene er pseudonymisert, er det fortsatt en personopplysning i lovens forstand, i motsetning til anonymiserte personopplysninger.

I personvernforordningens artikkel 25 om innebygget personvern blir pseudonymisering nevnt som et eksempel på egnede tekniske og organisatoriske tiltak for beskyttelse av personopplysninger. Der behandling for arkivformål i allmennhetens interesse, for formål knyttet til vitenskapelig eller historisk forskning eller for statistiske formål i henhold til artikkel 89 nr. 1 blir også pseudonymisering nevnt som en mulig garanti for å sikre at det er innført tekniske og organisatoriske tiltak.

I fortalen til personvernforordningen står det imidlertid at personopplysninger som er blitt pseudonymisert og som kan knyttes til en fysisk person ved hjelp av tilleggsopplysninger, bør ansees som personopplysninger.[55] Med at det står *bør* og ikke *skal* kan tolkes dit hen at det i tilfeller kan være at pseudonymiserte opplysninger ikke er personopplysninger i lovens forstand. Introduksjonen av begrepet pseudonym i personvernforordningen blir diskutert i artikkelen *Are 'pseudonymised' data always personal data?*[56] og på grunn av fortalepunkt 26 diskuteres det om pseudonymisering av personopplysninger i enkelte tilfeller kan sees på som anonyme opplysninger. I fortale 26 står det videre at:

" (...)Når det skal fastslås om en fysisk person er identifiserbar, bør det tas hensyn til alle midler som det med rimelighet kan tenkes at den behandlingsansvarlige eller en annen person kan ta i bruk for å identifisere vedkommende direkte eller indirekte, f.eks. utpeking. For å fastslå om midler med rimelighet kan tenkes å bli tatt bruk for å

[54] Personvernforordningen artikkel 4 nr. 5.
[55] Personvernforordningens fortale 26.
[56] Mourby et al., 2018, *Are 'pseudonymised' data always personal data?*

> identifisere den fysiske personen bør det tas hensyn til alle objektive faktorer, f.eks. kostnadene for og tiden som er nødvendig for å foreta identifikasjonen, idet det tas hensyn til teknologien som er tilgjengelig på behandlingstidspunktet, samt den teknologiske utvikling (…)"

Om en fysisk person er identifiserbar er altså avhengig av midler en med rimelighet kan tenkes benyttet, og rimelighet er blant annet et kostnadsspørsmål. I tillegg legger fortale 26 opp til at den behandlingsansvarlige tar hensyn til den teknologiske utviklingen. Nettopp dette kan sies å være svært utfordrende.

Samtidig står det beskrevet i personvernforordningen, at det å pseudonymisere personopplysninger er et sikkerhetstiltak ved behandling av personopplysninger. Dette ser vi eksempler på i artikkel 25 og 89 nevnt tidligere, samt fortalepunkt 28 ved at det identifiserende elementet er endret på og dermed beskyttet i større grad enn uten bruk av pseudonymisering. I tillegg bekrefter personvernforordningens fortale 78 at pseudonymisering kan være med å underbygge prinsippet om dataminimering.

Annen lovgivning beskriver pseudonymisering annerledes enn personvernforordningens definisjon. For eksempel defineres `pseudonyme helseopplysninger` i Forskrift om pseudonymt register for individbasert helse- og omsorgsstatistikk (IPLOS-registeret) slik:

> " (…) Helseopplysninger der identitet er kryptert eller skjult på annet vis, men likevel individualisert slik at det lar seg gjøre å følge hver person uten at identiteten røpes; definisjonen i personvernforordningen artikkel 4 nr. 5 gjelder ikke bestemmelsene om pseudonyme opplysninger i denne forskriften;" [57]

Selv om definisjonen er ulik, da IPLOS-registeret definerer pseudonymisering som en form for kryptering, som jeg omtaler i kapittel 3.3.4, er lovtekstene omforent med at pseudonymisering av en personopplysning fortsatt er en personopplysning og ikke en anonym opplysning.

Pseudonymisering kan muliggjøre personentydig behandling av personopplysninger. I tilfeller der det er behov for å følge et individ over tid, vil pseudonymisering være et sikkerhetstiltak for å beskytte det identifiserende elementet. For eksempel ved folkehelseundersøkelser vil

[57] Forskrift om IPLOS-registeret § 1-2 nr. 1.

bruk av pseudonymisering være en nyttig beskyttelsesmekanisme da samme pseudonym følger samme individ. Eksempler på dette er HUNT-Helseundersøkelsen i Nord-Trøndelag[58] og IPLOS-registeret nevnt over. Avidentifisering som ble omtalt i kapittel 3.3.2, vil ikke være et egnet sikkerhetstiltak i tilfeller der en skal følge personer over tid, slik som pseudonymisering er.

3.3.4 Kryptering

Ordet kryptografi kommer fra gresk og betyr hemmelig eller gjemt skrift. Kryptering er en betegnelse på et tiltak som benyttes for å gjøre opplysninger utilgjengelige for uvedkommende. I et personopplysningsøyemed er derfor kryptering et sikkerhetstiltak for å sikre konfidensialitet av personopplysninger. Alt kan krypteres, på forskjellige nivåer. For eksempel kan man kryptere en fil for å sørge for at uvedkommende ikke får tilgang til informasjonen i filen, man kan kryptere det identifiserende elementet i et sett av personopplysninger eller man kan kryptere en allerede pseudonymisert personopplysning for å heve sikkerheten rundt behandling av personopplysningene, for å nevne noen av bruksområdene til denne type sikkerhetstiltak.

Kryptering er et teknisk tiltak som omformer data slik at uvedkommende ikke kan forstå innholdet i dataene. Dette innebærer at informasjonen i dataene er gjort uforståelig i kryptert form. Dataene gir ikke mening for den som tilegner seg dem og dataene er derfor ikke `informasjon`. Et eksempel på kryptering av informasjon er krypteringsmaskinen Enigma som tyskerne benyttet seg av for å skjule informasjonen for de allierte ved kommunikasjon. Det er ikke bare stater og virksomheter som benytter kryptering for hemmelighold, tvert imot er det i dag flere kommersielle tjenester som tilbyr kommunikasjon via en kryptert kanal. Facebook-eide WhatsApp er et eksempel på en slik tjeneste.

For å få tilgang til informasjonen bak dataene må dataene dekrypteres, ved å dekryptere knekkes koden, og informasjonen som var beskyttet tilgjengeliggjøres. For å dekryptere data behøver man i utgangspunktet en nøkkel. I situasjoner der kryptering blir brukt til kommunikasjon mellom to parter vil mottaker av krypterte data også motta nøkkel for å kunne dekryptere dataene og se informasjonen. Grunnen til at krypterte personopplysninger fortsatt er personopplysninger i lovens forstand er at nøkkelen til informasjonen finnes. Det er

[58] NTNU: https://www.ntnu.no/hunt

altså en risiko for at en person kan identifiseres. Det finnes tre måter å få tilgang til kryptert data på.[59]

- Krypteringsnøkkelen er kjent (enten lovlig eller på annen måte)
- Knekke koden til nøkkelen eller algoritmen bak krypteringen ved hjelp av kryptoanalyse
- Kjenne til en teknisk-organisatorisk bakdør til innholdet

Ved et høyere beskyttelsesbehov kreves det en lengre nøkkel. Enkelt forklart kan en si at jo flere bits det er i nøkkelen jo vanskeligere er det å knekke koden. Vurderinger om lengden på nøkkelen sammen med krypteringsteknikk bør vurderes nærmere i en risikovurdering for å finne ut av hva som kan aksepteres av risiko.[60]

Det finnes to hovedtyper inndeling av krypteringsmekanismer. Symmetrisk og asymmetrisk kryptering.[61] Ved symmetrisk kryptering skjer krypteringen og dekrypteringen ved bruk av samme nøkkel mens med asymmetrisk kryptering er nøklene for kryptering og dekryptering ulike. Ved asymmetrisk kryptering vil det innebære et nøkkelpar der den nøkkelen som er kryptert er offentlig, mens nøkkel for å dekryptere er hemmelig og gis kun ut til de med autorisasjon. Asymmetrisk kryptering muliggjør derfor elektroniske signaturer. En elektronisk melding kan signeres ved at avsenderen krypterer med hemmelig nøkkel og en verifisering av signaturen kan skje ved av mottaker benytter den offentlige nøkkelen som tilhører avsenderen. Nøkkeladministrering ved elektronisk signatur tilbys i dag av sertifikatutbydere og omtales for PKI. Public Key Infrastruktures. I Norge er Buypass[62] og Commfides[63] eksempler på slike sertifikattilbydere. Sertifikatutsteder som for eksempel Buypass kan dermed bekrefte identiteten til den som sertifikatet er utstedt til.

Ovenfor er kryptering nevnt som et sikkerhetstiltak for å sikre konfidensialiteten til personopplysningene. Asymmetrisk kryptering sikrer også ivaretagelse av integriteten og sikrer autentisering av meldinger, dvs. at avsenderen av en melding faktisk er den personen

[59] Schartum og Bygrave, 2016, s. 307.
[60] For eksempel risikovurdering etter personvernforordningens artikkel 32 som går på personopplysningssikkerheten.
[61] NSM: https://www.nsm.stat.no/globalassets/dokumenter/veiledninger/systemteknisk-sikkerhet/u-02-grunnleggende-_tiltak-for-sikring-av-e-post---endelig.pdf
[62] Buypass: https://www.buypass.no/produkter/elektroniskID
[63] Commfides: https://www.commfides.com/commfides-virksomhetssertifikat/#e-idportalen

han/hun utgir seg for.[64] I tillegg kan asymmetrisk kryptering sørge for "ikke-benekting"[65] som innebærer muligheter for å knytte innholdet til avsenderen slik at hun ikke kan nekte for å stå bak det.[66]

Jeg skal ikke gå nærmere inn på krypteringsteknikker i denne oppgaven. Men hvis en behandlingsansvarlig velger å kryptere må det gjøres en konkret vurdering. I personvernøyemed kreves det for eksempel høyere sikkerhet rundt behandling av særlige kategorier av personopplysninger.[67]

3.3.5 Anonymisering

Ordet anonym stammer fra gresk og betyr `uten navn`. Hvis navnet er den eneste identifikatoren til en person, vil det ikke være mulig å identifisere et enkeltindivid. De resterende opplysningene vil da være anonyme og derfor ikke en personopplysning.

I personvernforordningen nevnes ikke anonymisering direkte. Det følger naturlig av at en anonym opplysning ikke er en personopplysning. Opplysningene som er anonyme er utenfor personvernforordningens virkeområde og reglene som for eksempel omfatter den registrertes rettigheter er ikke gjeldende. Det er altså ikke lenger en personopplysning og begrepet "den registrerte" eksisterer ikke.

Artikkel 29-gruppen beskrev anonyme opplysninger som: "Opplysninger som tidligere var knyttet til en person, der identifisering ikke lenger er mulig". De nevner videre at identifiserbarhetsvurderingen skal gjøres konkret i ethvert tilfelle og nevner spesifikt at man i statistiske data skal se på aggregeringen av opplysningene. Altså om størrelsen på utvalget er slik at det ikke skal være mulig å identifisere enkeltpersoner.[68]

I siste setning til fortale 162 i personvernforordningen står det at: "(…) Det statistiske formålet innebærer at resultatet av behandlingen for statistiske formål ikke er personopplysninger, men aggregerte data, og at dette resultatet eller personopplysningene

[64] Schartum og Bygrave, 2016, s. 307.
[65] "Ikke-benekting" - Uavviselighet. En mekanisme som kan knyttes til elektronisk samhandling og som er slik at de deltakende partene ikke kan benekte å ha deltatt i deler av eller hele samhandlingen."
Begrepet er hentet fra NOU 2001:10 Vedlegg 4.
[66] NOU 2001:10 s. 9.
[67] Etter personvernforordningens artikkel 9.
[68] Artikkel 29-gruppen, Opinion 4/2007 on *the concept of personal data*, 2007, s 21.

ikke brukes som støtte for tiltak eller avgjørelser som gjelder en bestemt fysisk person." Med dette konkretiserer forordningen at statistikk ikke er personopplysninger.

Det forekommer flere misforståelser og forskjellige tolkninger rundt begrepene jeg skriver om i dette kapittelet. Et kjent uttrykk innenfor helsesektoren er uttrykket `anonymt på forskers hånd`. Det beskriver en situasjon hvor en forsker har tilgang til opplysninger, men ikke tilgang til kodelisten. `Anonymt på forskers hånd` er en form for pseudonymisering, avidentifisering eller kryptering av et eller flere identifiserende element, slik at denne forskeren ikke vet identiteten til hvem det forskers på.

På grunn av at begrepet anonym blir benyttet i slike sammenhenger er det åpenbart at det ikke er like enkelt å vite hva som er en personopplysning og hva som er en anonym opplysning. Selv om forskeren ikke vet hvilke opplysninger som tilhører hvilke personer foregår det fortsatt en behandling av personopplysninger, og behandlingen av personopplysningene må skje i tråd med personvernregelverket.

3.4 Avsluttende kommentar

I dette kapittelet har jeg skrevet om begrepet personopplysning og identifisert flere teknikker en behandlingsansvarlig kan benytte seg av for å beskytte eller anonymisere personopplysninger. Imidlertid er det lite konkret omtalt i personvernforordningen da personvernforordningen fremsetter en risikobasert tilnærming til behandling av personopplysninger. Det er i stor grad opp til den behandlingsansvarlige å tolke regelverket og fastsette sikkerhetsnivå på opplysningene såfremt det ikke er nærmere definert i særlovgivning.

Resten av oppgaven er basert på casestudie jeg gjorde av SSB. Jeg har forsøkt å finne ut av hvordan de tolker regelverket de er underlagt både gjennom personopplysningsloven og statistikkloven. Til slutt vil jeg forsøke få svar på forskningsspørsmål nr. 3: *Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym?*

4 Statistisk sentralbyrå og statistikk

4.1 Innledning

I innledningen til NOU 2018:7 *Ny lov om offisiell statistikk og Statistisk sentralbyrå* står det om statistikksystemet i Norge at:

"Det er vanskelig å tenke seg et moderne samfunn uten god statistikk. Statistikk danner grunnlag for politiske beslutninger, understøtter en opplyst samfunnsdebatt og bidrar til felles forståelse av økonomiske og sosiale forhold. Statistikk bidrar med andre ord til velstand og velferd. Systemet har sitt tyngdepunkt i Statistisk sentralbyrå (...)"[69]

SSB er en faglig uavhengig institusjon som er ansvarlig for å samle inn, produsere og publisere offisiell statistikk relatert til økonomi, befolkning og samfunnet på nasjonalt, regionalt og lokalt nivå.[70] SSB driver også med forsking og analyse som skal bidra til ny kunnskap om økonomisk atferd og økonomiske virkninger av blant annet politiske tiltak.[71] Videre skal kunnskapen bidra til forskersamfunnet, til en kvalitativt bedre statistikk, og den skal gi analyseverktøy og resultater til bruk for offentlige organer og allmennheten.[72]

Statistikkloven gir regler for Statistisk sentralbyrås organisasjon og virksomhet. I tillegg regulerer statistikkloven innsamling og bruk av opplysninger til statistiske formål. SSB har røtter tilbake til 1878[73] og er i dag det nasjonale organet for utarbeidelse av offisiell statistikk om Norge. SSB er et forvaltningsorgan administrativt underlagt Finansdepartementet.[74]

Gjennom EØS-avtalen er Norge en del av EUs statistikksamarbeid, og EU-rettsakter på statistikkområdet tas inn som forskrift til statistikkloven. En sentral rettsakt er Forordningen om europeisk statistikk, som utgjør det rettslige rammeverket for utvikling, utarbeiding og formidling av europeisk statistikk. Forordningen omtales også som den «europeiske

[69] NOU 2018:7 s. 9.
[70] SSB: https://www.ssb.no/omssb/om-oss/
[71] SSB: https://www.ssb.no/omssb/om-oss/vaar-virksomhet
[72] SSB: https://www.ssb.no/forskning/forskning-i-ssb
[73] NOU 1988:19.
[74] Intervju med informant A.

statistikkloven» og regulerer samarbeidet i ESS (the European Statistical System). Som en aktør i ESS revideres (peer-review) SSB av Eurostat, som er EUs statistikkmyndighet.[75]

4.2 Samfunnsoppdrag og organisering

SSB har et todelt samfunnsoppdrag der SSB både er Norges sentrale statistikkmyndighet som skal utvikle, utarbeide og formidle offisiell statistikk, samtidig som de skal drive med forskning og analyse.[76] Likevel er det SSBs primære oppgave å utarbeide offisiell statistikk som skal dekke de fleste samfunnsområder, herunder befolkning og levekår, kommunal, fylkeskommunal og statlig virksomhet, ressurser og miljø og økonomi.[77] Som statistikkmyndighet skal SSB sørge for at befolkningen i Norge kan debattere, planlegge og ta beslutninger på grunnlag av pålitelig statistisk informasjon.[78]

SSB benytter nettsiden www.ssb.no som formidlingskanal for statistikken. Der publiseres artikler, analyser og rapporter om statistikk. All statistikk SSB utarbeider tilgjengeliggjøres i statistikkbanken.[79] Statistikkbanken inneholder i dag rundt 6000 tabeller med tidsserier. SSB tilbyr enhver å gjøre egne tilpassede uttrekk fra statistikkbanken. Det tilbys også et API (Application Programming Interface) som innebærer at de med behov kan integrere statistikk fra statistikkbanken inn i egne systemer.[80]

Figuren under viser organisasjonskartet til SSB.

[75] Intervju med informant A.
[76] Jf statistikkloven § 4-1 første og andre ledd.
[77] NOU:2018:7 s. 23 punkt 3.3.1.
[78] SSB: https://www.ssb.no/omssb/om-oss/vaar-virksomhet
[79] Intervju med informant B.
[80] NOU:2018:7 s. 23 punkt 3.3.1.

Figur 4 Organisasjonskart SSB

Som organisasjonskartet i figuren over viser, er SSB underlagt et styre. Informant A beskriver dette som en rar konstruksjon fordi det ikke er et reelt styre, men det er et slags faglig råd. Styret kan ikke overprøve direktøren i faglige spørsmål, men de skal høres for eksempel ved ansettelse av administrerende direktør. Denne konstruksjonen har vært debattert i lang tid og i henhold til den europeiske statistikkloven skal administrerende direktør være suveren, mens den norske statistikkloven legger noen oppgaver til styret. SSB fikk noen anmerkninger fra Eurostat som reviderte dem, med anbefaling om å se på styrets rolle. Dagens statistikklov med styret var i gråsonen for å etterleve den europeiske statistikkloven.[81]

I forslag til ny statistikklov (Prop. 72 LS 2018-2019) kommer problematikken med organiseringen frem slik informant A beskriver det. I proposisjonen kommer det frem at det er uklarhet om styrets rolle og om styret er en del av Statistisk sentralbyrå og dermed omfattet av

[81] Intervju med informant A.

den faglige uavhengigheten, eller om styremedlemmene fungerer som departementets representanter. Det siste underbygges av at det er departementet som oppnevner styret.[82]

Informant A forteller at organiseringen slik den er i dag, med et styre innebærer blant annet at styret og SSB stort sett rapporterer til Finansdepartementet felles, men dersom styret har egne meninger rapporteres det i to kanaler for å underbygge uavhengigheten til SSB og administrerende direktør.

På statistikkområdet har SSB tre avdelinger med underlagte seksjoner. Enhver seksjon er ansvarlig for sin statistikk, kalt statistikkansvarlig. Totalt er det 11 seksjoner i SSB som utarbeider statistikk på ulike områder. Det vil potensielt forekomme en behandling av personopplysninger i alle de tre statistikkavdelingene, men avdeling for person- og sosialstatistikk skiller seg ut da dette i utgangspunktet er personopplysninger som skal bli til statistikk. Avdeling for person- og sosialstatistikk, til høyre i organisasjonskartet i figur 4, har fem underliggende seksjoner.[83]

4.3 Rettslig grunnlag for behandling av personopplysninger

All behandling av personopplysninger må ha et behandlingsgrunnlag. Dette fremkommer av personvernforordningens artikkel 5 og 6. I artikkel 6 nr. 1 bokstav c står det om behandlingens lovlighet at: "behandlingen er nødvendig for å oppfylle en rettslig forpliktelse som påhviler den behandlingsansvarlige." SSB sin rettslige forpliktelse er hjemlet i Statistikkloven § 2-2 (1):

> "Kongen kan ved forskrift eller enkeltvedtak pålegge enhver å gi de opplysninger som er nødvendige for utarbeidelse av offisiell statistikk, så langt lovbestemt taushetsplikt ikke er til hinder for dette"

Informant A beskriver at det i utgangspunktet er SSB som bestemmer hvilken statistikk som skal lages, men at de er lydhøre for hva samfunnet etterspør. SSB skal speile samfunnet men får også føringer gjennom tildelingsbrev fra Finansdepartementet om hvilket tema som skal

[82] Prop. 72 LS 2018-2019 s. 78.
[83] SSB: https://www.ssb.no/omssb/om-oss/organisasjonskart/person-og-sosialstatistikk

belyses. SSB gjør vurderinger rundt kost-nytte verdien i forkant av innsamling av data ved å avveie belastning for den enkelte oppgavegiver sett opp mot nytteverdien i samfunnet.[84]

Informant A beskriver at det kun er lovbestemt taushetsplikt som er til hinder for at SSB kan få opplysningene. SSB er unntatt den generelle taushetsplikten etter forvaltningsloven § 13 bokstav b nr. 4 som sier at taushetsplikten i forvaltningsloven ikke er til hinder for at opplysninger kan leveres til SSB.[85]

I statistikkloven § 2-5 (1) fremkommer det at opplysningene hentet inn gjennom opplysningsplikten etter § 2-2, eller som er gitt frivillig, kun kan benyttes til utarbeidelse av offisiell statistikk.

Enkelte av rettighetene til den registrerte etter personvernforordningens tredje kapittel gjelder ikke for SSB. Det fremkommer av personopplysningsloven § 17 at retten til innsyn etter personvernforordningen artikkel 15 ikke gjelder for behandling av personopplysninger for statistiske formål, i samsvar med personvernforordningen artikkel 89 nr. 1. Det samme gjelder retten til retting og begrensning av behandling etter personvernforordningen artikkel 16 og 18. SSB informerer selv på nettsidene sine at de er unntatt innsynsretten og at den enkelte dermed ikke har krav på å få vite hvilke personopplysninger SSB behandler.[86]

4.4 Innsamling og lagring av personopplysninger hos SSB

SSB bruker begrepet oppgavegiver om den som bidrar til statistikken. Oppgavegiver kan for eksempel være en virksomhet, en kommune, en arbeidsgiver eller en person.[87] Det meste av SSBs personstatistikk hentes fra administrative registre og spørreundersøkelser. Hvis det er opplysninger som SSB har behov for som ikke finnes i et administrativt register, samles opplysningene inn via elektronisk rapportering, via Altinn. Altinn er primærkanal fra næringslivet til det offentlige. Enhver som er pliktet til å rapportere inn til SSB etter statistikkloven § 2-2 gjør dette hovedsakelig via Altinn, hvis opplysningene ikke innhentes

[84]Informant A: I forslag til ny lov skal offisiell statistikk omtales i egen nasjonalt statistikkprogram som vedtas av Kongen i statsråd.
[85] Intervju med informant A.
[86] SSB: https://www.ssb.no/omssb/personvern/personvern-og-datasikkerhet
[87] Erfaringer etter intervju med informant A.

direkte via systemene som nevnt over[88]. Informant A forteller at overføring av opplysninger skjer på samme måte som når en virksomhet skal rapportere sykefravær til NAV eller skatteopplysninger til skatt. I Altinn ligger det mange skjema som SSB har lagt klar, tilpasset den enkelte rapportering.[89] Det er også lagt til rette for filuttrekk fra diverse fagsystemer. Informant A forteller videre at SSB også har en filsluseløsning, der enkelte kan rapportere igjennom.

Figuren under skal illustrere prosessflyten ved elektronisk innrapportering til SSB.

Figur 5 Prosessflyt fra oppgavegiver til SSB

Ikonene til venstre i figur 5 skal illustrere person, næringsliv og offentlig sektor. Rapporteringen foregår ved å sende informasjonen til SSB via Altinn. SSB lagrer opplysningene på Kongsvinger.

SSB innhenter også personopplysninger gjennom intervjuer, enten via telefon eller ved å oppsøke folk hjemme.[90] Det er ikke mulig å reservere seg mot å delta i SSB sine statistiske undersøkelser, verken for næringslivet eller for privatpersoner. SSB begrunner oppgaveplikten, etter statistikkloven § 2-2, ved at undersøkelsene anses å ha stor samfunnsnyttig verdi at det vil si at det er lovpålagt å delta i undersøkelsen.[91] I tilfeller der opplysninger fra intervjuer skal sammenstilles med personopplysninger SSB allerede har hentet fra administrative registre (for eksempel opplysninger fra utdannelsesorganisasjoner

[88] Intervju med informant A.
[89] Intervju med informant A.
[90] SSB: https://www.ssb.no/omssb/om-oss/vaar-virksomhet
[91] SSB: https://www.ssb.no/omssb/personvern/dine-rettigheter/ingen-reservasjonsmulighet-mot-ssb

eller skatteopplysninger) kreves det behandlingsgrunnlag i form av samtykke[92] til denne sammenstillingen. Dette samtykket blir da innhentet i tilknytning til deltakelse i intervjuet.[93]

I utgangspunktet skiller ikke SSB på type opplysning som samles inn ved overføring fra Altinn eller fagsystem. Der det er en overføring av personopplysninger overtar SSB behandlingsansvaret for kopien av personopplysningene i det øyeblikket overføringen har skjedd. Etter at SSB har mottatt personopplysningene lagres disse i kjelleren til SSB på Kongsvinger. SSB benytter seg ikke av ekstern databehandler, men behandler personopplysningene selv.[94]

> *"Vi har opplysninger om det meste. Vi er nok den sivil instans i Norge som har samlet mest av opplysninger fordi vi får kopi av de fleste offentlige registre bortsett fra det som går på rikets sikkerhet og ikke helseregistrene."*[95]

Hos SSB vil det potensielt foreligge en behandling av personopplysninger både i næringsstatistikk og utenriksstatistikk.[96] I kapittel 3.1 kommer det frem at opplysninger om juridiske enheter, som for eksempel en bedrift eller kommune, i enkelte tilfeller kan si noe om enkeltpersoner, da er det snakk om en personopplysning. For ren person- og sosialstatistikk benytter SSB seg av fødselsnummer som en unik identifikator for å kunne skille personer fra hverandre. Det fremkommer i statistikkloven § 2-8 at SSB kan innhente folkeregisteropplysninger om det er nødvendig for utarbeidelse av offisiell statistikk. På samme måte som at taushetsplikten etter forvaltningsloven § 13 bokstav b nr. 4 nevnt i kapittel 4.3 ikke er til hinder for at opplysninger kan leveres ut til SSB, er heller ikke taushetsbelagte folkeopplysninger til hinder for at de kan leveres ut til SSB. I Lov om folkeregistrering (folkeregisterloven) er et av formålene at: "(...) Folkeregisteret skal kunne brukes til myndighetsoppgaver og offentlig forvaltning, forskning, statistikk og til å ivareta grunnleggende samfunnsbehov".[97]

Nevnt i kapittel 3.3.1 om entydig identifiserende personopplysninger setter personopplysningsloven begrensninger i bruk av fødselsnummer. Hvorvidt fødselsnummer

[92] Samtykke som behandlingsgrunnlag er nevnt i kapittel 3.2 under personvernprinsippet: "Lovlig, rettferdig og åpenhet".
[93] Intervju informant B.
[94] Intervju med informant A.
[95] Sitat fra informant A.
[96] Se figur 4: Organisasjonskartet til SSB.
[97] Folkeregisterloven § 1-2.

tillates brukt i behandling av personopplysninger avhenger av om den behandlingsansvarlige har saklig behov for slik bruk. I NOU 2001:10 *Uten penn og blekk - Bruk av digitale signaturer i elektronisk samhandling med og i forvaltningen* står det om bruk av fødselsnummer at den behandlingsansvarlige bør minst svare ja på følgende to spørsmål før fødselsnumre kan benyttes:

- Er det i behandlingen behov for entydig identifikasjon av en enkelt person?

- Skal opplysningene lovlig koples med personopplysninger i andre behandlinger?[98]

Informant A forteller at SSB er avhengig av en unik identifikator som gjør at de kan følge unike personer over tid. SSB får opplysninger fra mange ulike kilder og fødselsnummeret benyttes til å koble opplysningene til riktig person. Informant A forteller videre at den unike identifikatoren ikke nødvendigvis må være fødselsnummeret. Allikevel benytter de fødselsnummeret fordi de kan ha behov for dette dersom de må gå tilbake til kilden som en revisjonsaktivitet, eller hvis de oppdager feil som krever feilsøking.

Informant A eksemplifiserer en eventuell feil som om de oppdager at en 104-åring har tatt en nasjonal prøve. I et slikt tilfelle må de kunne gå tilbake til kilden for å finne ut av feilen og at de da har behov for en unik identifikator. Informant A forteller at for øvrig statistikkarbeid, trenger de ikke vite hvem en person er. De har behov for fødselsnummer for å utføre revisjonsarbeid.[99]

4.4.1 Dataminimering, integritet og fortrolighet

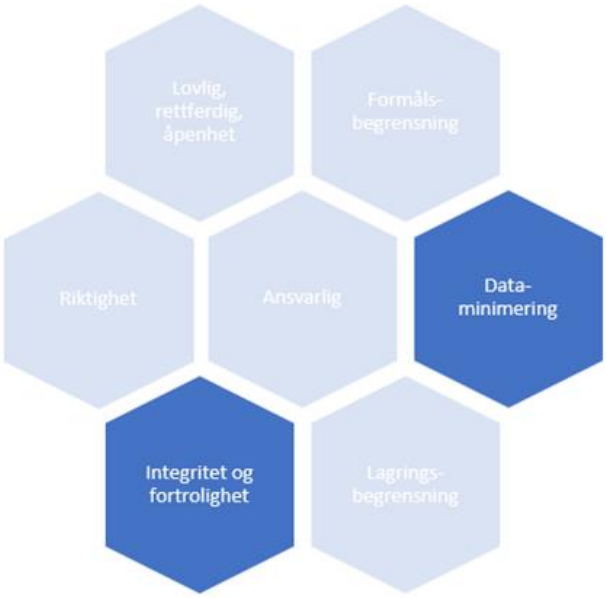
I kapittel 3.2 omtalte jeg personvernprinsippene generelt og hvordan behandling av personopplysninger for statistiske formål ikke anses som uforenlig med de opprinnelige formålene, etter personvernforordningens artikkel 5 bokstav b. Dette bekreftet informant A under intervjuet og forklarte om opplysningsplikten som fremkommer av statistikkloven § 2-2.

Om prinsippet om lagringsbegrensning forteller informant A forteller at et av kravene til offisiell statistikk er at statistikken som beskrives skal følges over tid, enten det er årlig, halvårlig eller korttidsstatistikk, og det skal presenteres en sammenliknende tidsserie. SSB har

[98] NOU 2001:10 s. 49.
[99] Intervju med informant A.

behov for historiske data for å bygge opp sammenlignende statistikk.[100] Denne informasjonen jeg fikk av informant A kan indikere at prinsippet om lagringsbegrensning blir hensyntatt, men fordi statistikken skal følges over tid, har de behov for historikken, og personopplysningene kan derfor ikke slettes etter utarbeidelse av en enkel statistikk. Dette er i henhold til unntaket for lagringsbegrensningen når formålet er statistikk, som nevnt i kapittel 3.2.

Figuren under skal illustrere de prinsippene som er mest relevante for statistikk og SSB: Prinsippet om "**dataminimering**" og prinsippet om " **integritet og fortrolighet**".



Figur 6 Skjult/uthevet figur av personvernprinsippene

SSB gjør en kost-nytte vurdering i forkant av innsamling av all data, også personopplysninger, og avgjør hva som er nødvendig av personopplysninger for å utarbeide statistikken. I denne vurderingen ser de på belastning for den enkelte oppgavegiver opp mot nytteverdien for samfunnet.[101] Informant A forteller at de ikke samler inn mer data enn strengt nødvendig og forteller videre at:

> *"Men det hender noen ganger at vi skal ha uttrekk fra fagsystem, og så sier de at det blir så mye plunder og heft at vi heller dumper over hele systemet hos dere så får dere ta det dere trenger"*[102]

[100] Intervju med informant A.
[101] Intervju med informant A.
[102] Sitat informant A.

SSB er da forpliktet til å slette det de ikke trenger og informant A forteller at de er bevisste på dette og ikke lagrer opplysninger de ikke har behov for.[103]

Informant A forteller om prinsippet "integritet og fortrolighet" ved å nevne informasjonssikkerheten deres som innebærer blant annet: sikre systemer, rutiner og retningslinjer for behandling av opplysninger og taushetsplikt for enhver som har eller har hatt stilling i eller oppdrag for SSB. Videre forteller han mer detaljert at for å sikre etterlevelse av personvernforordningen med forsvarlig behandling av personopplysninger lagrer SSB personopplysninger i ulike databaser. Her skjelnes mellom flate filer og relasjonsdatabaser. Personopplysningene er derfor ikke samlet på ett sted og blir ikke sammenstilt før aggregering. De ulike filene og databasene er tilgangsstyrt basert på tjenstlig behov.

En flat fil forteller et situasjonsbilde, mens relasjonsdatabasene fyller på data hele tiden så man kan følge med på alt som har skjedd med den enkelte person fra ulike tidspunkt. For eksempel vil relasjonsdatabasen gi informasjon fra du fikk den første utdanningsopplysningen din fra barneskolen til du er ferdig doktorand. SSB følger hver enkelt person over tid og sted i den samme databasen.[104] Informant A forteller at de har tre slike databaser; FDtrygd, nasjonal utdanningsdatabase og Edag (inntekt/arbeidsforhold). I databasene blir fødselsnummer erstattet med et pseudonym i form av et statistikknummer. For de flate filene har SSB nå et pseudonymiseringsprosjekt som går på å pseudonymisere fødselsnummeret også i de flate filene. Dette for å tilfredsstille krav etter personvernforordningen, i tillegg til at det er lagt inn forslag i den nye statistikkloven om et eksplisitt krav om å pseudonymisere.[105] Informant A forteller at dette skulle egentlig vært på plass i et moderniserings- og digitaliseringsprogram i forkant av personvernforordningen, men forteller at programmet ble kraftig nedjustert og forsinket. SSB jobber med å finne en ad hoc-løsning. Det er de flate filene som blir benyttet til utarbeidelse av statistikk mens relasjonsdatabasene blir brukt til forskningsformål.

[103] Intervju med informant A.
[104] Intervju med informant A.
[105] Intervju med informant A.

5 Konfidensialitet i statistikken

5.1 Rettslige krav til konfidensialitet

Definisjonen av statistikk blir omtalt i statistikkloven § 1-2 (1): *"Statistikk er tallfestede opplysninger om en gruppe eller fenomen, som fremkommer ved sammenstilling og bearbeiding av opplysninger om de enkelte enhetene i gruppene eller et utvalg av disse enhetene, eller ved systematisk observasjon av fenomenet"*.

Videre blir offisiell statistikk omtalt i statistikkloven § 1-2 (2): *"Statistikk som gjøres tilgjengelig for allmennheten av Statistisk Sentralbyrå eller annet statlig organ"*.

Når SSB utarbeider statistikk, aggregeres/grupperes enkeltobservasjoner slik at opplysningene havner på gruppenivå, som presenteres i tabellform. Deretter må SSB sikre at tabellene anonymiseres før de skal publiseres, dette følger av statistikkloven § 2-6:

> "Opplysninger hentet inn etter fastsatt opplysningsplikt, eller som er gitt frivillig, skal ikke i noe fall offentliggjøres slik at de kan føres tilbake til oppgavegiver eller annen identifiserbar enkeltperson til skade for denne, eller til urimelig skade når det gjelder selskaper med begrenset ansvar, kommandittselskaper og andre sammenslutninger, stiftelser og offentlige organer og virksomheter."

Informant A forteller at det er statistikkloven § 2-6 som regulerer anonymiteten til statistikken de publiserer og forteller videre at de pålagt av § 2-6 til å publisere statistikk som i utgangspunktet er helt anonym.[106] Som det fremkommer av lovteksten nevnes ikke anonymt eksplisitt, det står derimot at opplysninger SSB offentliggjør ikke skal være til skade for oppgavegiver. Og som nevnt i kapittel 4.4 kan oppgavegiver være både enkeltpersoner og virksomheter. Likevel er tolkningen hos SSB at etter § 2-6 skal statistikken som hovedregel være anonym.[107]

[106] Intervju med informant A.
[107] Unntak blir omtalt i kapittel 5.3.

Krav til konfidensialitet i statistikken gjenspeiler seg i norske lovverk, EU-reguleringer og internasjonale prinsipper. SSB er underlagt EUs retningslinjer for statistikk, og FNs prinsipper for offisiell statistikk.

Retningslinjene til EU baserer seg på 16 prinsipper der det for hvert prinsipp fremstilles indikatorer for god praksis for å kunne evaluere gjennomføringen av retningslinjene. Prinsipp nummer fem i EUs retningslinjer omhandler statistisk konfidensialitet og informasjonssikkerhet, hvor det står:

> *"Statistical Confidentiality and Data Protection: The privacy of data providers, the confidentiality of the information they provide, its use only for statistical purposes and the security of the data are absolutely guaranteed."* [108]

FNs prinsipper for offisiell statistikk har ti prinsipper der prinsipp nummer seks sier at:

> *"Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."* [109]

Konfidensialiteten i statistikken gjelder uavhengig av om den som bidrar til statistikken (oppgavegiver) er en person eller en virksomhet (fysisk eller juridisk person). Der det er snakk om fysisk person vil personvernforordningen være gjeldende og SSB må blant annet behandle personopplysningene på en måte som sikrer tilstrekkelig sikkerhet for personopplysningene i henhold til prinsippet om integritet og fortrolighet etter artikkel 5 nr. 1 bokstav f.

Selv om hovedregelen er at all statistikk er underlagt de samme kravene kommer det frem i intervjuene med begge informantene at de statistikkansvarlig er tydelige på at konsekvensene ved å publisere opplysninger som viser seg å ikke være anonym har ulike konsekvenser, og de har satt terskelverdier deretter. Dette omtaler jeg mer om i kapittel 5.3.

[108] EU: https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7
[109] FN: https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf

5.2 Avsløring og identifisering

Det ligger i definisjonen av statistikk at dette er tallfestede opplysninger om en gruppe eller et fenomen, og dermed ikke personopplysninger. SSB benytter seg i liten grad av begrepet anonymt når det er snakk om statistikk, de benytter begrepet *avsløring*.[110] Begrepet avsløring ble definert av statistikeren Tore Dalenius i Statistisk tidsskrift i 1977.[111]

En avsløring har funnet sted hvis man gjennom statistikk eller statistiske data lærer noe mer om en statistisk enhet enn man kunne vite uten statistikken eller de statistiske dataene.[112]

En avsløring er ikke det samme som å identifisere noen i et datasett. Informant A forteller at statistikken SSB publiserer skal være anonym og forteller videre at det ikke skal være mulig å avsløre informasjon om noen. Ingen skal identifiseres i statistikken til SSB, men om en statistikkbruker klarer å identifisere noen, eller seg selv, så er det fordi denne personen har informasjon fra før som gjør at den for eksempel kan se at dette må være en gitt person. Det som er avgjørende er at statistikken ikke skal avsløre informasjon om noen. Du skal ikke få vite noe mer om statistiske enheter inkludert individer via SSB sin statistikk.[113] I statistikk over mindre geografiske områder vil lokalkunnskap hos statistikkbrukerne lettere bidra til identifisering.

For å eksemplifisere hvordan en identifisering kan forekomme i en statistikk forklarer informant A et eksempel om en som har deltatt i en spørreundersøkelse og SSB lager statistikk på bakgrunn av denne undersøkelsen. Det er en tabell som sier at det er så og så mange personer som har de og de egenskapene. Tenk deg at tabellen viser at det bare er én person i undersøkelsen som har svart katt. Hvis jeg vet at du var med i den undersøkelsen og vet at du har svart katt, klarer jeg å skjønne at det dette er deg, men statistikken avslørte ikke noe mer informasjon om deg. Hvis det sto at den ene i undersøkelsen som hadde svart katt, også hadde noen andre egenskaper, så kunne statistikken avslørt de andre egenskapene hvis jeg visste du hadde svart katt. For eksempel ved at statistikken avslørte hvilken utdannelse du har eller hvilken inntekt du har, det ville vært strengt forbudt.[114]

[110] Erfaring etter intervju med begge informantene samt intern dokumentasjon fra SSB.
[111] Dalenius, 1977.
[112] SSBs interne oversettelse av definisjonen på avsløring fra Dalenius. Fra internt dokument.
[113] Intervju med informant A.
[114] Intervju med informant A.

I forslag til ny lov om offisiell statistikk og Statistisk sentralbyrå[115] blir begrepet avsløring introdusert i § 7: "Statistisk konfidensialitet ved formidling av offisiell statistikk". SSB spilte inn ønske om å erstatte «*identifisere den statistiske enheten*» med «*avsløre informasjon om den statistiske enheten*» i høringsrunden. SSB forklarer i høringen at de mener denne terminologien er mer presis og i tråd med internasjonale formuleringer og begreper som brukes innen offisiell statistikk.[116]

5.2.1 Hvordan kan avsløring finne sted?

Statistikk blir presentert gjennom aggregerte tabeller og SSB skiller mellom spesielt to ulike tabeller der en avsløring kan finne sted. Frekvenstabeller og mengdetabeller. En frekvenstabell forteller hvor mange enheter det er i en tabellcelle, det er en opptelling av antall enheter i hver celle. Et eksempel på en frekvenstabell vil være en tabell som viser antall personer i ulike aldersintervaller og deres sivilstatus. En mengdetabell er en tabell der antall enheter i cellen er erstattet av en sum. For eksempel kan en mengdetabell vise en gjennomsnittslønn. Bak en mengdetabell vil det forekomme en frekvenstabell. I figur 7 og 8 under vises eksempler på de ulike tabellene.

[115] Prop. 72 LS (2018 – 2019).
[116] Prop. 72 LS (2018 – 2019) s. 26.

Antall personer	Alder			
Sivilstatus	**18-29**	**30-59**	**60+**	**SUM**
Ugift	13	7	1	**21**
Gift	10	19	11	**40**
Skilt	2	9	15	**26**
SUM	**25**	**35**	**27**	**87**

Figur 7 Tabell 1A eksempel på frekvenstabell

Inntekt, 1000 kr.	Alder			
Sivilstatus	**18-29**	**30-59**	**60+**	**SUM**
Ugift	4 423	10 529	450	**15 402**
Gift	3 792	8 392	4 291	**16 475**
Skilt	978	5 568	6 081	**12 627**
SUM	**9 193**	**24 489**	**10 822**	**44 504**

Figur 8 Tabell 1B eksempel på mengdetabell

Figur 7 er et eksempel på en frekvenstabell der tallet i cellen viser antall enheter som er med i underlaget. Figur 8 er et eksempel på en mengdetabell der cellene viser inntekt. Bak figur 8 foreligger figur 7. Dette innebærer at i figur i cellen [Ugift] [30-59] i figur 8 er det syv personer med i underlaget.

Man kan også tenke seg tabeller der cellene inneholder andre størrelser enn totaler/summer for en numerisk variabel, for eksempel gjennomsnitt eller medianer. Dette er mindre vanlig i SSBs statistikkproduksjon. For personstatistikk er frekvenstabeller mest utbredt mens mengdetabeller er mest utbredt i økonomisk statistikk. [117] Tabell 1B i figur 8 er et eksempel på en personstatistikk i en mengdetabell.

SSB har definert ulike scenarier der avsløring kan finne sted. Sannsynligheten for avsløring er typisk høyere jo mindre tabellpopulasjonen er. Scenariene der avsløring kan finne sted er i `små tall` og `dominans`.

[117] Intervju med informant B.

Små tall

Små frekvenser i tabellceller, for eksempel 1, 2 eller 3[118], er ikke uten videre avslørende. Man lærer ikke noe mer enn det man måtte vite for å kunne plassere enhetene i riktig tabellcelle. Dersom SSB ikke bare publiserer antall enheter i cellene, men også summerer verdier knyttet til en numerisk variabel knyttet til enhetene, for eksempel en inntekt eller et stønadsbeløp, vil verdien på denne variabelen knyttet til enheten i tabellcellen uten videre avsløres.[119]

Figur 9 under er et eksempel på en frekvenstabell med `små tall`.

Antall personer	Alder			
Sivilstatus	**18-29**	**30-59**	**60+**	**SUM**
Ugift	13	7	**1**	**21**
Gift	10	19	11	**40**
Skilt	**2**	9	15	**26**
SUM	**25**	**35**	**27**	**87**

Figur 9 Frekvenstabell med små antall

I figur 9 er cellene med `små tall` uthevet med rød farge. Dette gjelder cellene med verdiene [Skilt] [18-29] og [Ugift] [60+].

Om statistikkbruker vet om noen som er med i tabellgrunnlaget, for eksempel med å allerede kjenne til deres sivilstatus og alder kan statistikkbrukeren plassere dem i riktig tabellcelle. Men i tabellen (figur 9) lærer en ikke noe mer om personen og det vil ikke finne sted noen avsløring. Hvis frekvenstabellen i figur 9 krysses med en ny variabel, derimot, for eksempel en utdanningsvariabel eller lønnsvariabel, vil utdanningen eller lønnen til den ugifte 60+ åringen bli avslørt.

Hvis en mengdetabell som i vist i figur 8 skal publiseres og det bak mengdetabellen foreligger en frekvenstabell (figur 7) vil dette potensielt føre til en avsløring. Avsløring vil forekomme om statistikkbruker vet hvem som er med i statistikkunderlaget med alder [60+] med sivilstatus [Ugift].

Figur 10 under viser et eksempel på en mengdetabell med avsløring.

[118] Tallene 1,2 eller 3 er ikke endelig eller definerte terskler hos de statistikkansvarlige i SSB men er her illustrert som et eksempel.
[119] Intervju med informant B.

Inntekt, 1000 kr.	Alder			
Sivilstatus	18-29	30-59	60+	SUM
Ugift	4 423	10 529	450	15 402
Gift	3 792	8 392	4 291	16 475
Skilt	978	5 568	6 081	12 627
SUM	9 193	24 489	10 822	44 504

Figur 10 Mengdetabell med avsløring 1

I figur 10 har det forekommet en avsløring da statistikkbruker på bakgrunn av informasjon den sitter på fra før har identifisert en person i en celle. Denne personens inntekt blir da avslørt. Inntekten på 450.000 kr er uthevet i rødt.

Hvis vi går tilbake til frekvenstabellen med `små tall` i figur 7 ser man at det er mulig for enda en avsløring i denne tabellen. I cellen [Skilt] [18-29] kan det være tilfelle der statistikkbruker vet hvem én av de to personene er. Der denne frekvenstabellen krysses med en mengdetabell, og hvis statistikkbruker vet hvem en av de to er, kan statistikkbruker for eksempel trekke slutning at en ikke har tjent mer enn 978.000 kr. Figuren 11 under viser en mengdetabell der to celler er identifisert som potensielt avslørende.

Inntekt, 1000 kr.	Alder			
Sivilstatus	18-29	30-59	60+	SUM
Ugift	4 423	10 529	450	15 402
Gift	3 792	8 392	4 291	16 475
Skilt	978	5 568	6 081	12 627
SUM	9 193	24 489	10 822	44 504

Figur 11 Mengdetabell med avsløring 2

I figur 11 over er cellene [Skilt] [18-29] og [Ugift] [60+] markert i rødt da det er fare for avsløring.

Om statistikkbruker selv er en del av tabellcellen [Skilt] og [18-29] kan denne personen regne ut inntekten til den andre ved å trekke fra sin egen inntekt. Dersom de to i cellen var gift med hverandre, vil de begge få avslørt sin inntekt ovenfor sin fraskilte ektefelle. Informant B forteller at i en nasjonal tabell med bare to personer i en celle vil det ofte være usannsynlig at en slik avsløring kan finne sted, men i en tabell for en liten kommune eller grunnkrets vil det kunne skje.

Dominans

En avsløring kan forekomme i mengdetabeller ved at en enhet gir et mye større bidrag til celletotalen enn de andre i mengdetabellen. SSB de kaller dette scenariet `dominansregelen`. For å forhindre `dominans` har SSB laget to regler som sier at:

- De største bidragene til en celle totalt ikke må overstige en viss prosent av totalen i cellen, eller
- De to største bidragene ikke må overstige en viss prosent av totalen i cellen.[120]

`Dominans` forekommer som oftest i statistikk med økonomiske enheter (foretak, bedrifter) men kan forekomme også i personstatistikk, for eksempel inntektsstatistikk.[121] I tillegg kan det forekomme en avsløring i KOSTRA statistikk (kommune/stat rapportering).[122] I KOSTRA statistikk får SSB inn regnskap fra kommunene. Denne type statistikk inneholder mye informasjon om blant annet barnevern og sosialtjenester. I disse statistikkene kan det ligge informasjon som ikke absolutt ikke skal avsløres, forteller informant B. Videre forteller han at en avsløring her potensielt kan si noe om hvor mye en kommune bruker av utgifter på ett barn. Et eksempel på slik informasjon er om en kommune har hatt utgifter på barnevern eller sosialhjelp, og om kommunene har få innbyggere. Det er derfor viktig at det ikke kan tilbakeføres hvor store utgifter en kommune har på et barn, for eksempel om statistikkbruker kjenner til et barnehjemsbarn i kommunen.[123]

I tabellene 1A og 1B i figur 7 og 8 har cellene [Ugift] [30-59] syv personer med samlet inntekt på 10.529.000 kr. For å eksemplifisere hvordan `dominans` kan oppstå har jeg i figur 12 under laget en tabell der de de syv personene er sortert etter synkende inntekter. Inntektene er deretter aggregert nedover.

[120] De to prosentene holdes utenfor offentligheten for ikke å undergrave bruken av regelen, men informant B gav meg denne prosentdelen muntlig under intervjuet.
[121] Intervju med informant B.
[122] Hentet fra ssb.no: KOSTRA (KOmmune-STat-RApportering) er et nasjonalt informasjonssystem som gir styringsinformasjon om kommunal virksomhet. Opplysningene hentet inn av Statistisk sentralbyrå nyttes til utarbeidelse av styringsinformasjon for kommuner, bydeler og fylkeskommuner.
[123] Intervju med informant B.

i	y_i		
1	8 529		100 %
2	2 000		
3	…		…
4	…		…
5	…		…
6	…		…
7	…		…
Sum	10 529	100	

Figur 12 Eksempel på avsløring igjennom `dominans`

I tabellen i figur 12 ser vi at personen med høyest inntekt tjener 8.529.000, mens personen med nest høyest inntekt tjener 2.000.000. Det innebærer at de to personene til sammen tjener 100% av totalen i de to cellene. En kan se for seg at statistikken i figur 12 ville avslørt ovenfor den med høyest inntekt at de andre i underlaget hadde svært lav eller ingen inntekt. For eksempel så kan personen med høyest inntekt ha en ansatt som er med i statistikkunderlaget. Det er rimelig å tro at personen med høyest inntekt vet lønnen til sin ansatte. I et slik tilfelle vil personen med høyest inntekt vite at de resterende fem (om personen vet at disse er med i underlaget) ikke har lønn. I et eksempel som dette ville SSB ikke publisert statistikken, men benyttet seg av teknikker for å forhindre avsløring, hvilket skrives om i kapittel 6.

Knyttet opp mot forskningsspørsmål 3: *Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym?* SSB har her vurdert at det finnes to scenarier for hvordan en avsløring kan oppstå. Gjennom `små tall` og `dominans`. De statistikkansvarlige hos SSB har fastsatt regler i form av terskelverdier som brukes for å unngå avsløring. I kapittel 6 beskrives hvilke teknikker SSB benytter seg av for å unngå avsløring i statistikken.

5.3 Konfidensialitetsutvalget i SSB

I kapittel 5.1 omtalte jeg statistikkloven § 2-6 og hvordan SSB tolker denne bestemmelsen. SSB har som hovedregel at statistikk som offentliggjøres ikke skal kunne føres tilbake til en

statistisk enhet eller annen identifiserbar enkeltperson på tross av at lovteksten i § 2-6 gir muligheter for dette så lenge det ikke er til *skade* for oppgavegiver.[124]

Informant A forteller at unntaksvis, av hensyn til hensiktsmessig oppbygging av statistikken, kan være nødvendig at opplysninger publiseres slik at enkelte oppgavegivere eller andre blir identifisert. I slike tilfeller er det et absolutt krav at en slik identifisering ikke er til skade for fysisk person eller til urimelig skade for juridisk person, slik det fremkommer av § 2-6. Det er den statistikkansvarlige seksjonen som skal gjøre disse vurderingene. Om statistikkansvarlige mener at et unntak er på sin plass skal den endelige avgjørelsen tas av administrerende direktør, etter at saken er behandlet av Konfidensialitetsutvalget hos SSB.

Som jeg nevnte innledningsvis i delkapittel 2.3.1 sitter begge informantene i Konfidensialitetsutvalget, et internt utvalg i SSB som skal bidra med rådgivning på konfidensialitetsspørsmål og vurdere om en statistikk skal publiseres selv om det er en fare for avsløring. Informant A leder dette utvalget.

Informant A beskriver prosessen slik:

- Dersom den statistikkansvarlige enheten kommer over en tabell som viser at denne ikke er helt anonym, må de bruke en teknikk for å lage den anonym.[125]

- Hvis statistikkansvarlig mener det er meningsløst å publisere denne og mener at SSB må gjøre unntak, skal de fremme saken i linjen via sin direktør til Konfidensialitetsutvalget.

- Den ansvarlige seksjonen må beskrive statistikken og planlagt detaljeringsnivå og begrunnelse for detaljeringsnivået. Det skal også begrunnes hvorfor statistikkansvarlig ser behov for å publisere den aktuelle statistikken på et slikt detaljert nivå at det vil være mulig å identifisere individuelle statistiske enheter. Den statistikkansvarlige skal selv vurdere risikoen for mulig avsløring av enkeltenheter da det er den statistikkansvarlige som best kjenner statistikken og dataunderlaget. Her skjelner SSB mellom fysisk person og juridisk person og deretter om den juridiske personen er privat eller offentlig.

[124] Eller til urimelig skade når det gjelder selskaper med begrenset ansvar, kommandittselskaper og andre sammenslutninger, stiftelser og offentlige organer og virksomheter jf. statistikkloven § 2-6, tredje ledd.
[125] Mer om anonymiseringsteknikker i kapittel 6.

- Vurderingen sendes deretter til Konfidensialitetsutvalget. Utvalget består av medlemmer fra alle fagenhetene.

- Konfidensialitetsutvalget diskuterer om vilkårene er tilstede. Vilkårene er `hensiktsmessighet` og at avsløringen ikke skal være til `skade`. Hvis første vilkår (hensiktsmessighet) er tilstede, vurderer de det andre vilkåret, om dette vil være til skade.

- Hvis Konfidensialitetsutvalget er enig med statistikkansvarlig og mener at det ikke er til skade, så gir de anbefaling til direktøren som tar den endelige beslutningen. Om Konfidensialitetsutvalget mener at avsløringen kan være til skade så stopper utvalget publiseringen.[126]

Her gjøres det ulike vurderinger ut i fra om opplysningene er om fysiske personer eller juridiske personer, og tilsvarende om opplysningene gjelder offentlige eller private rettssubjekter. Det kan være vanskelig å foreta en slik vurdering forteller informant A. Hva som vil være til `skade` avhenger dels av opplysningenes karakter, dels av hva slags statistisk enhet opplysningene beskriver.[127]

I mange av sakene avslår utvalget søknaden, og viser til at statistikken kan publiseres anonymt på en hensiktsmessig måte ved å bruke teknikker for å forhindre avsløring. Informant A forteller at det er få saker som går til direktøren: ca. 30 saker på 15 år. Videre forteller han at offentlige enheter generelt må tåle mer enn kommersielle og eksemplifiserer dette med da Postverket hadde monopol og måtte finne seg i at det kunne være statistikk som avslørte informasjon om virksomheten til Postverket. [128]

På spørsmål om de tar hensyn til om den potensielt avslørende informasjonen allerede er offentlig tilgjengelig svarer informant A at Konfidensialitetsutvalget i utgangspunktet ikke tar hensyn til om informasjonen allerede er offentlig tilgjengelig. Informant A beskriver videre at statistikkloven sier at opplysningene de har samlet inn, er underlagt taushetsplikt. Når SSB produserer statistikk skal de i utgangspunktet håndtere det som konfidensielt.

[126] Intervju med informant A.
[127] Intervju med informant A.
[128] Intervju med informant A.

Hvis opplysningene er tilgjengelige via andre offentlige kilder vil det ikke være urimelig skade, så noen ganger vil det kunne være med i vurderingen, forteller informant A. Men i vurderingen om publisering av statistikk kan være til skade, og når de selv har sagt at de ikke har noe imot at statistikken publiseres, så vil det gjøre Konfidensialitetsutvalget sin `skadevurdering` enklere.

I forslag til ny statistikklov er det tatt inn åpning for å publisere statistikk som kan være avslørende basert på samtykke, eller dersom det som avsløres allerede er offentlig kjent.[129] Dette betyr at SSB kan vurdere å innhente samtykke fra oppgavegiver om oppgavegiver samtykker til at den type informasjon kan bli avslørt via statistikken.

[129] Kommentarer i referat fra intervjuet med informant A.

6 Anonymisering i statistikken

6.1 Metoder for å forhindre avsløring i statistikken

I veilederen til Artikkel 29-gruppen: *Anonymisation Techniques* fra 2014, deles anonymiseringsteknikker i to grupper, randomisering og generalisering. Det poengteres til stadighet at ingen av anonymiseringsteknikkene er uten svakheter og teknikkene gjerne kan brukes i kombinasjon med hverandre for å styrke anonymiseringen.[130]

Aggregering inngår i anonymiseringsteknikken generalisering og innebærer at opplysninger om enkeltindivider er slått sammen eller at opplysningene er kombinert med andre individer slik at ikke opplysningene kan spores tilbake til et individ. Opplysninger som gis karakteriserer en masse og ikke enkelte statistiske enheter. Definisjonen av statistikk etter statistikkloven blir derfor å se på som en aggregering av opplysninger, i tillegg til at opplysningene er tallfestet.[131]

Seksjonene i SSB som er ansvarlig for statistikken sin har til dels hatt ulike verktøy, alt fra manuelt arbeid, til hjemmesnekrede løsninger og hyllevare. Informant B forteller at de forsøker få til størst mulig grad av standardisering og at de har fattet interesse for standardisering spesielt i senere tid. Informant B forteller videre at fordi det er mye mer detaljert statistikk nå enn før er standardisering viktig. SSB er derfor avhengig av å redusere arbeidsbyrden ved å manuelt ta i bruk teknikker de benytter for å unngå avsløring i statistikken. Informant B forteller at spesielt i store tabeller tar det mye tid å forhindre avsløring og forteller videre at det er enormt mye tid å spare og at kvaliteten blir bedre. I tillegg unngår de menneskelige feil. Som en del av profesjonaliseringsarbeidet benytter SSB nå en programvare som heter Tau-Argus. Programvaren er utviklet i Nederland, og finansiert av Eurostat. SSB kan selv sette terskelverdier i programvaren og integrere disse i annen programvare.[132]

[130] Artikkel 29-gruppen: Opinion 05/2014 on *Anonymisation Techniques*
[131] Statistikkloven § 1-2 (1).
[132] Intervju med informant B.

For å sørge for at statistikken i tabellform er anonym og ikke avslører noe informasjon om en person eller andre statistiske enheter benytter SSB seg av tre ulike teknikker for å forhindre dette. Disse er undertrykking/prikking, avrunding og ytterligere aggregering.

6.1.1 Undertrykking/prikking

Undertrykking eller "prikking" innebærer at tallet ikke offentliggjøres, men erstattes med et kolon. SSB benytter seg av begge begrepene. Jeg vil heretter benytte meg av begrepet prikking.

Ved bruk av prikking finnes det en risiko for at statistikkbruker kan regne seg tilbake til verdien av cellen på grunnlag av summen i en tabell. I slike tilfeller er det behov for å prikke flere celler. SSB kaller dette sekundærprikking. Informant B beskriver at med store tabeller er det svært komplisert å vite hva som skal sekundærprikkes. På spørsmål om SSB i enkelte statistikker unngår å ha med summen, slik at sekundærprikking ikke er nødvendig svarer informant B at statistikken vil miste verdien sin.[133] Der prikking er benyttet, blir verdien i den cellen som kan føre til avsløring erstattet med et kolon. (:)

Hvis tabell 1A (figur 7 i kapittel 5.2.1) skulle publiseres, og teknikken for å forhindre avsløring var prikking, ville den muligens sett ut som tabellen under (figur 13).

Antall personer	Alder			
Sivilstatus	**18-29**	**30-59**	**60+**	**SUM**
Ugift	:	:	:	**21**
Gift	10	19	11	**40**
Skilt	:	9	:	**26**
SUM	**25**	**35**	**27**	**87**

Figur 13 Tabell med undertrykking/prikking

I cellen med alder [60+] og sivilstatus [Ugift] var det kun en person med i underlaget, derfor blir denne cellen prikket. I cellen med alder [18-29] og sivilstatus [Skilt] var det kun to personer med i underlaget, denne blir også prikket. For å unngå at statistikkbruker kan regne ut hvilket tall som skulle være i den prikkede cellen (ved å ta summen og trekke fra tallene i de resterende cellene) kreves det sekundærprikking i denne tabellen. I eksempelet i figur 13 er også cellen med alder [18-29] og sivilstand [Ugift] sekundærprikket. Det samme er cellen

[133] Intervju med informant B.

med alder [60+] og sivilstatus [Skilt]. Det vil ikke være mulig for en statistikkbruker og se hvilke celler som er prikket og hvilke som er sekundærprikket. De fire prikkede cellene skal forhindre avsløring i `små tall` nevnt i kapittel 5.2.1. I figur 13 er en femte celle prikket. Cellen med alder [30-59] og sivilstatus [Ugift] er prikket for å forhindre avsløring i `dominans` også omtales i kapittel 5.2.1.

For å teste SSBs prikking slik det ser ut i produksjon, gikk jeg inn i statistikkbanken på SSB sine nettsider og gjorde en spørring på Utsira kommune. Utsira er Norges minste kommune[134] og dermed var det en større sannsynlighet for at `små tall` vil forekomme. I tillegg valgte jeg den yngste aldersgruppen, som i mindre grad eier egen bolig.

Jeg valgte statistikken "Ligningsverdi" ved å søke på:

→[statistikkvariabel]: Personer med ligningsverdi

→[region]: Utsira

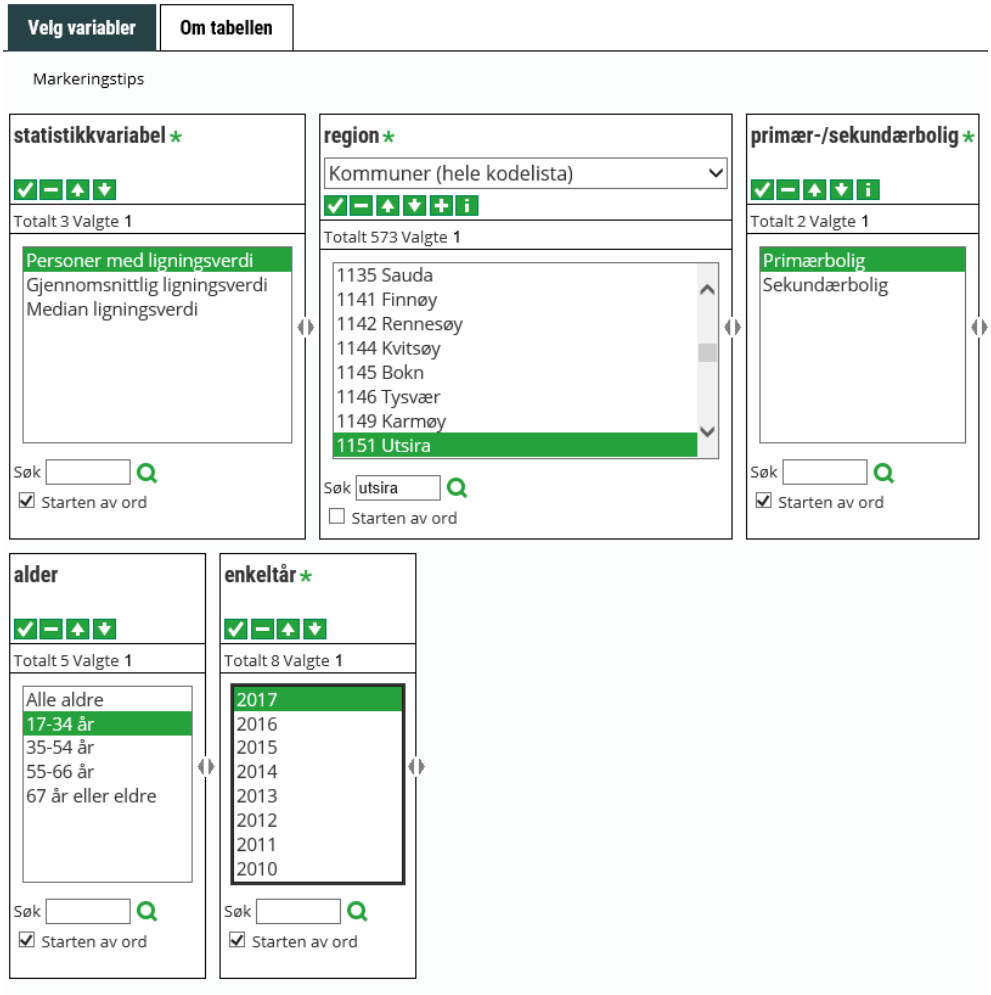
→[primær-/sekundærbolig]: Primærbolig

→[alder]: 17-34

→[enkeltår] 2017

Figuren under er et skjermbilde fra spørringen jeg utførte i statistikkbanken til SSB.

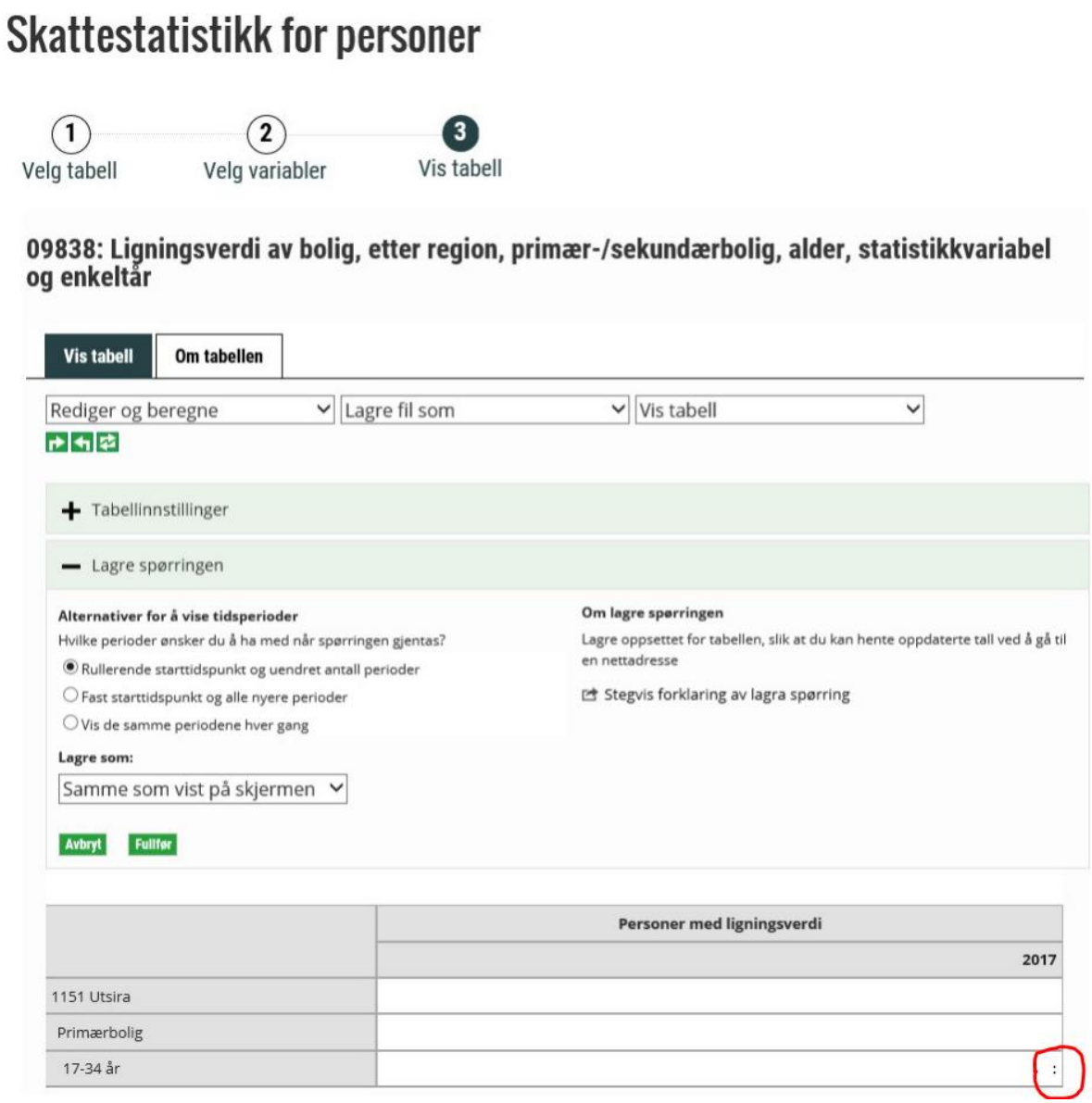
[134] Utsira hadde 1 januar 2019 196 innbyggere. Informasjonen er hentet i statistikkbanken til SSB.



Figur 14 Spørring i statistikkbanken

Valg av kommune (liten kommune) og aldersgruppe (unge mennesker som i mindre grad eier egen bolig) skulle øke sannsynligheten for at resultatet av spørringen ble som jeg håpet, to prikker: (:)

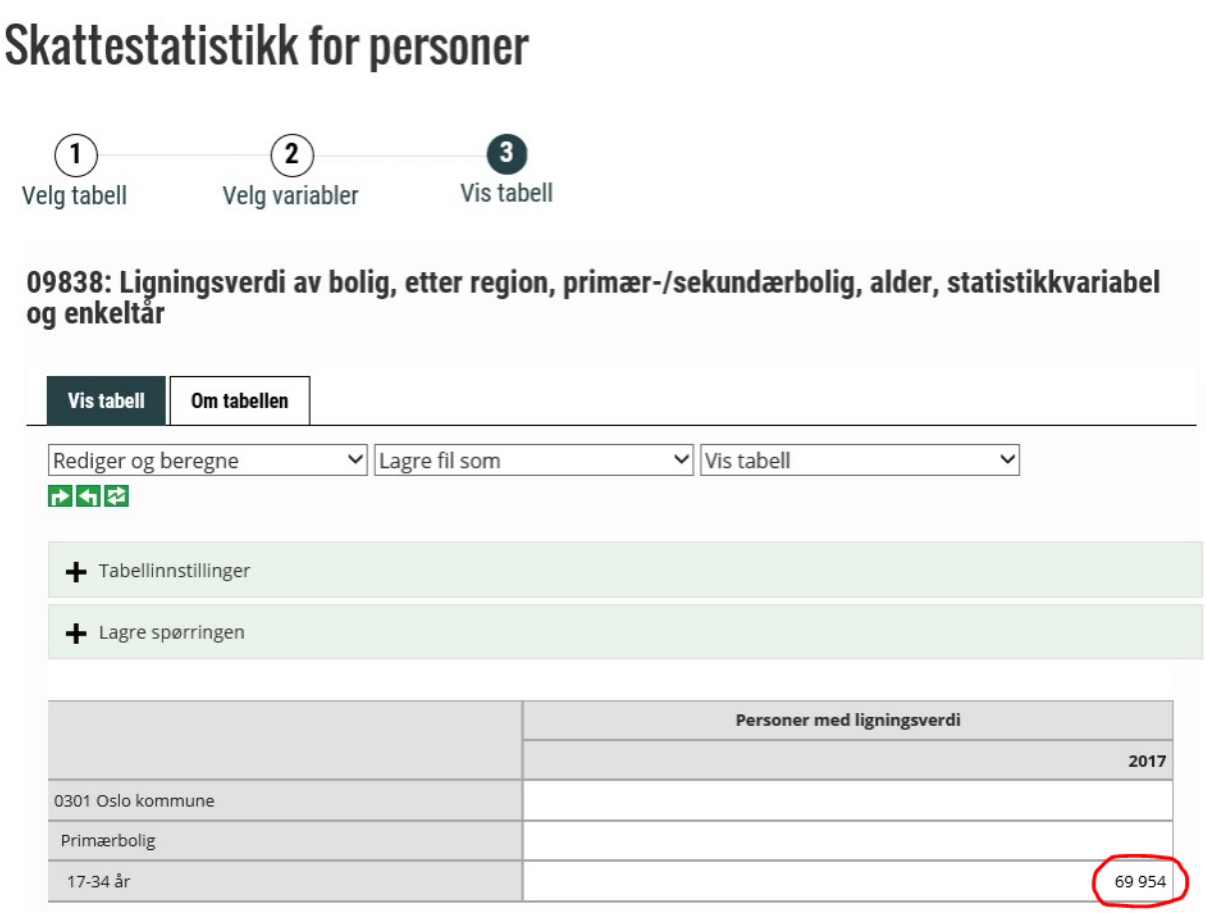
Figur 15 under viser resultatet av spørringen:



Figur 15 Statistikkbanken: Resultater med prikking

I figur 15 ser vi at verdien på antall personer med ligningsverdi er prikket da det er fare for avsløring.

For å se et eksempel på lignende statistikk hvor det ikke er fare for avsløring utførte jeg samme spørring på Oslo kommune. I figur 16 under er resultatet av spørringen.



Figur 16 Statistikkbanken: Resultater uten prikking

I figur 16 er det ikke fare for avsløring og resultatet kan vise verdien.

6.1.2 Avrunding

Avrunding innebærer at tall avrundes til et multiplum av et basetall, oftest 3, men det kan også være 4 eller 5. Denne metoden brukes oftest helst på `små tall` i frekvenstabeller, men kan i noen tilfeller også brukes i mengdetabeller.[135]

Avrunding av tall benyttes når det er få personer eller statistisk enheter i hver tabellcelle, som skrevet om i kapittel 5.2.1. `Små tall` kan for eksempel 1 eller 2, eller hvis statistikkbruker kan finne ut noe mer enn det som skal til for å plassere en person eller statistisk enhet i riktig tabellcelle. Informant B beskriver at flere av de statistikkansvarlige enhetene velger å beskytte med å avrunde og prikke mer enn det som er strengt tatt nødvendig. Dette for å ta høyde for at de på et senere tidspunkt lager ny tabell med flere variabler med mengde eller volumtabell.[136]

Ved bruk av avrunding har SSB satt hovedregel om at terskelverdien må være minst tre personer eller statistiske enheter i en tabellcelle. Men informant B poengterer at det er den statistiske ansvarlige enhet som bestemmer denne terskelverdien. Den statistikkansvarlige enheten med tabeller av mer sensitiv karakter som for eksempel barnevern, der setter enheten ofte en høyere terskelverdi, foreller informant B. Også innenfor kriminalitetsstatistikken har informant B erfaring med at statistikkansvarlig har vært restriktive og overbeskyttet tabellene. Erfaringene informant B har gjort seg er at i tilfeller der de statistikkansvarlige har vært usikre på hva som kan være avslørende blir de strenge selv. Det innebærer at de statistikkansvarlige prikker eller avrunder mer slik at de unngår å fremme saken til Konfidensialitetsutvalget.[137]

De samme vurderingene blir benyttet enten metoden for å hindre avsløring er prikking eller avrunding.[138]

Et eksempel på hvordan avrunding blir brukt er at i stedet for 1 eller 2 står det 0 eller 3 i en frekvenstabell. Informant B beskriver at ofte er det 0-tallet som lager problemer. Mye av hensikten med avrunding er å skape usikkerhet om 0-er er virkelige nuller eller avrundede `små tall`. For eksempel i en tabell hvor en rad har bare 0-er utenom i en kolonne, så skal du ikke vite sikkert om alle 0-ene faktisk er 0. Avrunding er en form for støy som skal skape usikkerhet på om man kan trekke konklusjoner eller ikke. Fordelen med avrunding fremfor

[135] Intervju med informant B.
[136] Intervju med informant B.
[137] Se kapittel 5.3 om Konfidensialitetsutvalget i SSB.
[138] Intervju med informant B.

prikking, spesielt i frekvenstabell er hvis en statistikkbruker vil summere opp tallene. Det er ikke mulig å summere opp med en prikk. Men med avrunding vil statistikkbruker kunne summere med noe feilmargin. Av hensyn til noen typer statistikk er avrunding dermed bedre enn prikking.[139]

Hvis tabell 1A (figur 7 i kapittel 5.2.1) skulle publiseres, og teknikken for å forhindre avsløring var avrunding, ville den muligens sett ut som tabellen under.

Antall personer	Alder			
Sivilstatus	**18-29**	**30-59**	**60+**	**SUM**
Ugift	13	7	**0**	**20**
Gift	10	19	11	**40**
Skilt	**3**	9	15	**27**
SUM	**26**	**35**	**26**	**87**

Figur 17 Tabell med avrunding

I tabellen i figur 17 er cellen med sivilstatus [Skilt] og alder [18-29] og sivilstand [Ugift] og alder [60+] blitt avrundet med å erstatte opprinnelig tall med 3 og 0. Figur 17 viser kun et eksempel på avsløring i `små tall`, ikke i `dominans`.

Ved spørringer i statistikkbanken der avrunding er benyttet har SSB denne informasjonen som en fotnote:

(...) Alle ett-tall og to-tall i tabellen er endret til '0' eller '3' for å ivareta personvernet. Når tallene aggregeres til høyere regionalt nivå, vil summen kunne avvike noe fra det faktiske tallet.[140]

SSB viser med dette åpenhet om metoden de benytter for å forhindre eventuell avsløring, i tillegg til at de forklarer hvordan dette kan føre til eventuelle statistiske avvik.

6.1.3 Ytterligere aggregering

Denne metoden gjør grupperinger av variabler enda grovere ved å slå sammen kategorier. På denne måten vil ofte de strukturene som kan gi grunnlag for avsløring forsvinne, spesielt på `små tall`.[141] For eksempel kan SSB velge å ha større aldersspenn i en gruppe eller gå fra

[139] Intervju med informant B.
[140] Statistikkbanken: https://www.ssb.no/statbank/
[141] E-post fra informant B 01.04.2018.

bydel til kommunenivå for å unngå `små tall`, som nevnt i kapittel 5.2.1. Ytterligere aggregering blir ikke brukt av SSB på samme måte som prikking og avrunding fordi SSB lager statistikk, og den skal være sammenlignbar for eksempel på kommunenivå.[142] SSBs tabellpubliseringer har en utforming med inndelinger som er faste fra år til år. Når SSB da lager en statistikk som skal være sammenlignbar over hele landet på kommunenivå da kan de ikke slå sammen for eksempel Utsira, som kun har noen få hundre innbyggere, med en annen kommune, da dette gir feil statistikk.

Detaljeringsnivå på statistikken er hovedsakelig noe som gjøres når en ny statistikk skal utformes. Dette gjøres av de seksjonene som har ansvaret for statistikken. Detaljeringsnivået kan avhenge av i hvilken grad variablene som inngår oppfattes som «sensitive». En slik aggregering innebærer ofte et stort informasjonstap sammenlignet med avrunding eller prikking. Men hvis tabellene blir så «tynne» at veldig mye forsvinner i prikking eller nulles ut ved avrunding kan det å inndele i grovere kategorier være mere formålstjenstlig.[143]

[142] Intervju med informant A.
[143] Avsnittet om ytterligere aggregering er basert på erfaring fra intervjuene samt noen oppfølgingsspørsmål jeg hadde til informant B sendt på e-post den 05.04.2019.

7 Oppsummering og avsluttende kommentarer

Jeg startet oppgaven med å se nærmere på personopplysningsbegrepet. For å vite hvordan man kan anonymisere en personopplysning er det essensielt å kunne skille de to opplysningstypene. Her har jeg benyttet meg av juridisk metode for å finne frem til gjeldende rett og tolke regelverket. Etter en redegjørelse for personopplysningsbegrepet, etterfulgt av personvernprinsippene, beskrev jeg ulike grader av identifisering og hvordan de identifiserende elementene kan beskyttes ved hjelp av fire forskjellige teknikker. Avidentifisering, pseudonymisering, kryptering og til slutt anonymisering. De fire teknikkene kan alle bidra til å etterleve kravene i personvernforordningen og sørge for tilstrekkelig sikkerhet rundt behandlingen av personopplysninger.

I kapittel 3 ble forskningsspørsmål 1 og 2 besvart: *Hvilke rettsregler gjelder for identifisering og for beskyttelse av identiteter knyttet til personopplysninger? Hvilke teknikker kan benyttes for å beskytte de identifiserende elementene?*

Videre i oppgaven startet casestudie med SSB inn. Jeg har benyttet meg av samfunnsvitenskapelig metode og dybdeintervju for å se på hvordan SSB som behandlingsansvarlig anonymiserer statistikken der opplysningene stammer fra et enkeltindivid. Jeg fant ut at SSB benytter seg av entydig identifiserbare personopplysninger i form av fødselsnummer for å koble opplysninger til en identitet. Når SSB lager statistikk aggregeres/grupperes enkeltobservasjoner slik at opplysningene havner på gruppenivå, som presenteres i tabellform. Aggregering av personopplysninger er en anonymiseringsteknikk som benyttes i statistikken.

SSB har metoder og teknikker for at de aggregerte opplysningene ikke skal kunne avsløre informasjon om et enkeltindivid. De ansvarlige seksjonene har satt terskelverdier på antall enheter som må være med i en celle. Denne metoden omtaler de som `små tall`. I tillegg har de en metode for å unngå `dominans` som går ut på at få enheter gir et mye større bidrag til celletotalen enn de andre i mengdetabellen. SSB benytter seg av to ulike teknikker for å unngå avsløring. Disse er prikking og avrunding. SSB bruker en programvare fra Eurostat der ansvarlig seksjon selv definerer terskelverdier for sin statistikk. Unntaksvis publiserer SSB statistikk som kan avsløre informasjon om et enkeltindivid, da skal det ikke være til skade for

denne personen. SSB har organisert et Konfidensialitetsutvalg som skal gi råd i slike situasjoner. Der Konfidensialitetsutvalget og statistikkansvarlig mener at statistikken med fare for avsløring skal publiseres, går saken til administrerende direktør som tar den endelige avgjørelsen.

Jeg innledet denne oppgaven med påstanden om at anonymisering er viktigere og samtidig vanskeligere enn noen gang. SSB har jobbet med statistikk siden slutten av 1800-tallet og har dermed lang erfaring med anonymisering. På tross av dette finnes det muligheter for at informasjon man tilegner seg igjennom statistikken kan avsløre informasjon om en enkeltperson. SSB jobber aktivt med å unngå at en avsløring skal finne sted i deres statistikk. I kapittel 5 og 6 svarte jeg på forskningsspørsmål 3: *Ved utarbeidelse av statistikk fra personopplysninger: Hvordan sørger Statistisk sentralbyrå for at statistikken de utarbeider er og forblir anonym?*

Begge informantene virket selvsikre da de snakket om anonymitet i statistikken, eller forhindring av avsløring, som er begrepet SSB bruker. Inntrykket mitt etter intervjuene var at statistikk er noe SSB har jobbet med lenge og er svært dyktige på. På spørsmål om hvordan de sørger for at statistikken er anonym etter publisering, og om de utfører reidentifiseringstester, fikk jeg til svar at dette ikke gjøres. Det virker som at de ikke har et reelt behov for etterkontroll da de ikke har hatt avvik når det kommer til avsløring i statistikk. Begge informantene påpekte under intervjuene at SSB er avhengig av oppgavegiveres og publikums tillit til at opplysninger de gir fra seg ikke blir avslørt eller brukt til andre formål enn forskning og statistikk. Mangel på tillit kan i verste fall skade SSBs muligheter til å oppfylle sitt samfunnsoppdrag og slik være til skade for hele samfunnet. Kombinert med den økende trenden med ønske om en mer detaljert statistikk vil SSB måtte fortsette sitt arbeid med å forhindre avsløring i statistikken, enten om dette er nye metoder, strengere terskelverdier, andre verktøy eller økt etterkontroll i form av reidentifiseringstester.

Mangel på tillit kan forekomme uavhengig av statistikken de utarbeider. Dårlig håndtering av personopplysninger internt kan for eksempel føre til mistillit. I forslag til ny statistikklov (Prop. 72 LS 2018-2019) er det lagt frem eksplisitte krav til informasjonssikkerheten. Kravene innebærer blant annet logging og etterfølgende kontroll. Det blir spennende å se på hvordan SSB velger å håndheve disse kravene. Den nye statistikkloven muliggjør at SSB kan innhente helseopplysninger. Helseopplysninger går innunder særlige kategorier av personopplysninger og det kreves beskyttelse deretter.

7.1 Videre arbeid

Min erfaring er at statistikk er enklere å anonymisere da statistikk allerede er tallfestede, aggregerte opplysninger. Utlevering av mikrodata til forskere gjennom tjenesten microdata.no derimot, har helt andre utfordringer. Det hadde vært spennende å se nærmere på jobben SSB gjør her for å sørge for anonymitet. SSB har i samarbeid med Norsk senter for forskningsdata (NSD) laget en ny tjeneste nå som er i grenseland mellom det å låne ut til forskning og til å publisere statistikk. Den heter microdata.no. Før tjenesten microdata.no var hovedregelen at SSB utleverte data til forskere. Informant A fortalte at det ofte forekommer avvik, for eksempel ved at data kommer på avveie eller forskeren benytter dataene til andre formål. I begrepet data her inkluderer mulige personopplysninger. Tjenesten microdata.no vant Datatilsynets pris "Innebyggede personvern i praksis" i 2018. I begrunnelsen la juryen vekt på at microdata.no fjerner personvernrisikoen ved forskning på mikrodata. Det ble også påpekt at Norge er først i verden med en slik infrastruktur.[144]

En annen del av SSB som det hadde vært spennende å se nærmere på, er hvordan SSB gjør en kost-nytte vurdering i forbindelse med statistikken de utarbeider. Informantene forteller begge at de i vurderingen av ny statistikk ser på belastning for den enkelte oppgavegiver opp mot nytteverdien i samfunnet. I et personvernperspektiv hadde det vært interessant å få innblikk i en slik vurdering der oppgavegiver er en fysisk person. I henhold til et forprosjekt kalt *Utvikling av en forløpsdatabase på barnevern* har SSB kartlagt mulighetene for å organisere barnevernsdata med å utvide barnevernsstatistikken slik at det blir enklere å levere barnevernsdata til forskning og analyse på forløpsform, som er sammenlignbar over tid. Her ser vi et eksempel på at det er ønskelig med en mer detaljert statistikk enn tidligere, og slik informant B beskrev under intervjuet er dette en økende trend. Et innblikk i en kost-nytte vurdering, spesielt der oppgavegiver i dette tilfellet er barn, som krever et særlig vern i henhold til personvernforordningen,[145] hadde vært interessant. At oppgavegiver også er barnevernsbarn indikerer at personopplysningene er av en sensitiv karakter og krever muligens en vurdering av personvernkonsekvensene etter artikkel 25 i personvernforordningen.[146]

[144] Datatilsynet: https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-20192/microdata.no-vant-pris/
[145] Se spesielt personvernforordningens fortale 38.
[146] SSB: https://www.ssb.no/sosiale-forhold-og-kriminalitet/artikler-og-publikasjoner/utvikling-av-en-forlopsdatabase-pa-barnevernsomradet-forprosjekt

Litteraturliste

Litteratur

Blume, Peter (2006): *Juridisk metodelære.* 4 utgave. København: Jurist- og Økonomiforbundets Forlag.

Dalenius, Tore (1977): *Towards a methodology for statistical disclosure control*, Statistik Tidskrift 15 (429-444), 2—1.

Grønmo, Sigmund (2004): *Samfunnsvitenskapelige metoder*. Bergen: Fagbokforlaget.

Miranda Mourby et al. Miranda Mourby, Elaine Mackey, Mark Elliot, Heather Gowans, Susan E. Wallace, Jessica Bell, Hannah Smith, StergiosAidinlis, Jane Kaye. (2018): *Are 'pseudonymised' data always personal data?* Published by Elsevier Ltd.

Olsen, Thomas (2015): *Personvernøkende identitetsforvaltning.* Oslo: Senter for rettsinformatikk.

Schartum, Dag Wiese og Bygrave, Lee A (2016): *Personvern i informasjonssamfunnet. En innføring i vern av personopplysninger*. 3 utgave. Bergen: Fagbokforlaget.

Schartum, Dag Wiese i samarbeid med Bygrave, Lee A (2008)*: Utredning om fødselsnummer, fingeravtrykk og annen bruk av biometri i forbindelse med lov om behandling av personopplysninger § 12.* Rapport bestilt av Justis- og beredskapsdepartementet.

Rettskilder

Lov om offisiell statistikk og Statistisk Sentralbyrå (statistikkloven) LOV-1989-06-16-54

Lov om folkeregistrering (folkeregisterloven) LOV-2016-12-09-88

Lov om behandling av personopplysninger (personopplysningsloven) LOV-2018-06-15-38

Forskrift om innsamling og behandling av helseopplysninger i Kreftregisteret (Kreftregisterforskriften) FOR-2001-12-21-1477

Forskrift om overføring av biobankmateriale til utlandet FOR-2004-02-26-511

Forskrift om pseudonymt register for individbasert helse- og omsorgsstatistikk (Forskrift om IPLOS-registeret) FOR-2006-02-17-204

Forskrift om innsamling og behandling av helseopplysninger i Norsk pasientregister (Norsk pasientregisterforskriften) FOR-2007-12-07-1389

Forskrift om gjennomføring av EØS-rettsakter om europeisk statistikk (Forskrift om europeisk statistikk) FOR-2008-06-20-632

Prop. 72 LS (2018 – 2019) Proposisjon til Stortinget (forslag til lovvedtak og stortingsvedtak)

NOU 1988:19 *Lov om offisiell statistikk og Statistisk Sentralbyrå*

NOU 2001:10 *Uten penn og blekk — Bruk av digitale signaturer i elektronisk samhandling med og i forvaltningen*

NOU 2018:7 *Ny lov om offisiell statistikk og Statistisk sentralbyrå*

Domsavgjørelser

Judgment in Case C-582/14. Patrick Breyer v Bundesrepublik Deutschland

Elektroniske kilder

Brønnøysundregistrene (felles datakatalog):

https://fellesdatakatalog.brreg.no/ (lest:29.04.2019)

Buypass:

https://www.buypass.no/produkter/elektroniskID (lest: 21.10.2018)

Commfides:

https://www.commfides.com/commfides-virksomhetssertifikat/#e-idportalen (lest: 21.10.2018)

Datatilsynet:

https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/informasjonssikkerhet/hvordan-anonymisere-personopplysninger/ (lest: 19.03.2018)

https://www.datatilsynet.no/personvern-pa-ulike-omrader/internett-og-apper/id-tyveri/identitetstyveri--hva-trenger-id-tyven-og-hvordan-beskytter-du-deg---/ (lest: 06.07.2018)

https://www.datatilsynet.no/regelverk-og-verktoy/verktoy/ordbok-a-til-a/ (lest: 12.12.2018)

https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-20192/microdata.no-vant-pris/ (lest: 18.03.2018)

Direktoratet for forvaltning og IKT:

https://internkontroll-infosikkerhet.difi.no/risikostyring/risikovurdering (lest:05.08.2018)

EU:

https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7 (lest: 18.11.2018)

FN:

https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf (lest: 18.11.2018)

Nasjonal sikkerhetsmyndighet:

https://www.nsm.stat.no/globalassets/dokumenter/veiledninger/systemteknisk-sikkerhet/u-02-grunnleggende- tiltak-for-sikring-av-e-post---endelig.pdf (lest: 20.07.2018)

NTNU:

https://www.ntnu.no/hunt (lest: 02.02.2019)

Skatteetaten:

www.skatteetaten.no/person/folkeregister/fodsel-og-navnevalg/barn-fodt-i-norge/fodselsnummer/ (lest: 18.11.2018)

Statistisk sentralbyrå:

https://www.ssb.no/omssb/om-oss/vaar-virksomhet (lest: 02.01.2019)

https://www.ssb.no/omssb/om-oss/organisasjonskart/person-og-sosialstatistikk (lest: 01.03.2019)

https://www.ssb.no/forskning/forskning-i-ssb (lest: 08.09.2018)

Muntlige kilder og informanter

Intervju med informant A hos SSB sine kontorer i Oslo den 24.01.2019

Intervju med informant B hos SSB sine kontorer i Oslo den 07.02.2019

E-post med informant B den 01.04.2019

Andre kilder:

Artikkel 29-gruppen, opinion 4/2007 on *the concept of personal data* (2007)

Artikkel 29-gruppen, opinion 05/2014 on *Anonymisation Techniques* (2014)

Datatilsynets veileder: *Grunnleggende personvernprinsipper* (2018)

Datatilsynets veileder: *Sporing i det offentlige rom - Bruk av WiFi, Bluetooth, nettvarder (beacons) og intelligent videoanalyse* (2016)

SSB. Notat 2019/04: *Utvikling av en forløpsdatabase på barnevernsområdet – Forprosjekt*

Vedlegg: Intervjuguide

Intervjuguide Statistisk sentralbyrå

I forbindelse med masteroppgave i forvaltningsinformatikk ved UiO var formålet med det første intervjuet å få et innblikk i SSB generelt og personvern spesielt. Overordnet om hvordan de behandler personopplysninger, hvordan de anonymiserer og hvordan dere organisatorisk håndterer risiko.

Formålet med intervju nummer to er å få et innblikk i hvordan SSB anonymiserer statistikken de publiserer.

Som informant vil du kan være representert som informant X. Stillingstittelen din vil bli presentert i oppgaven

1) Samtykker du til lydopptak?

    a. Lydopptak vil kun benyttes av meg til å transkribere dette intervjuet

    b. Etter at formålet med behandlingen er over og intervjuet er transkribert så vil jeg slette lydopptaket.

I etterkant av intervjuet vil jeg sende ut referat til deg for godkjenning. Her har du mulighet til å kommentere, utdype eller presisere innholdet slik at det blir korrekt.

.

Spørsmål til Informant A: Organisering

1. Kan du beskrive litt om rollen din i SSB? *(herunder: beskrivelse, ansvar knyttet til rollen, rapporteringsvei)*

2. Kan du fortelle kort om samfunnsoppdraget SSB har? *Hvem er dere og hva gjør dere?*

3. Kan du fortelle overordnet hvordan SSB nå er organisert? *(herunder: hvordan er fagområder fordelt i organiseringen, enheter per fag? Hva med "sikkerhetsorganiseringen" og ansvarlige roller i organiseringen? Inkludert beslutningstakere i avgjørelser. Tolkningsansvarlige/ juridisk enhet? Har dere et personvernombud? Beredskap?*

    a. *Mer spesifikt om statistikk: hvordan er statistikken organisert?*

Spørsmål til informant A: Juridiske rammer

1. Kan du fortelle litt rundt hvilke juridiske rammer må dere forholde dere til når det kommer til utarbeidelse av statistikk? *(Herunder kan du fortelle litt om deres styrende dokumenter, relevante lovverk som personvernforordningen, statistikkloven, FNs prinsipper for offisiell statistikk, europeiske retningslinjer for statistikk, personvernforordningen. Er lovverkene harmoniserte?)*

2. Kan du enkelt beskrive hvordan dere innhenter data/personopplysninger til å omforme til statistikk? *(herunder: beskrivelse av datafangsten. Via altinn/ andre måter dere innhente data på? Skiller dere på type data ved innsamling for eksempel forskjell på næringsdata og rene personopplysninger?)*

3. Har dere måtte gjøre endringer i hvordan dere behandler personopplysninger etter den nye personvernforordningen? *(enderinger fra gammel til ny personopplysningslov sommeren 2018. for eksempel Garantier etter artikkel 89.*

Videre skal jeg diskutere med informant A hva jeg ønsker videre:

1) Innblikk i en anonymiseringsprosess der dere anonymiserer personopplysninger. For eksempel der et enkeltindivid blir en del av statistikk. Hvordan dette gjøres i praksis. Teknisk/metodisk/organisatorisk. Gjerne se på flere anonymiseringsteknikker (om SSB benytter flere)

2) Risiko knyttet til behandlingen og anonymiseringen. *(Herunder i hvilken grad dere utfører risikovurderinger knyttet til behandlingen? Gjør dere reidentifiseringstester. Eksterne eller interne tester? Utfører dere behandlinger der dere er pålagt å utføre en personvernkonsekvendutredning (DPIA) og i så fall hva gjør dere behandlinger dere anser kan ha høy risiko for den registrerte? Og noe om hvordan de håndterer avvik)*

Spørsmål til informant B: <u>Anonymisering av personopplysninger</u>

1. Hvordan praktiserer dere personvernprinsippene i henhold til personvernforordningen. *(herunder: spesielt dataminimering og integritet og fortrolighet. NB: Der det behandles personopplysninger. Ikke anonymiserte)*

2. Er dere behandlingsansvarlig eller databehandler? *(når forekommer eventuelt en behandlingsansvaret)*

3. Hvilken *type* opplysninger er en del av grunnlaget for utarbeidelse av statistikken du anonymiserer? *(herunder: Særlige kategorier? Andre sårbare grupper?)*

4. Mottar dere allerede anonyme opplysninger/statistikk eller anonymiserer dere alltid selv?

5. Hvis dere anonymiserer selv: Hvordan mottar dere personopplysninger av en behandlingsansvarlig. *(Datafangst via Altinn?)*

6. Er personopplysningene dere mottar knyttet til en identifikator? Hvilken*? (For eksempel: fødselsnummer?)*

7. Etter dere har mottatt personopplysningene. Hvordan lagres disse? *(herunder tilgangsstyring, Hvor lenge lagres de? Slettes personopplysningene eller anonymisering? Automatisk sletting? Har dere behov for personopplysningene etter anonymisering? Lagrer dere personopplysninger selv eller benytter dere databehandler)*

8. Hvilken teknikk benytter dere for å anonymisere?

9. Hvem hos dere utfører anonymiseringen. *(Hvor i organiseringen, hvem eller hva? Kan være et system som anonymiserer, hvem er ansvarlig?)*

10. Hvilke verktøy benyttes? *(laget verktøy selv? Hyllevare? Verktøy som benyttes av andre i EU?)*

11. Hvordan avgjør dere om opplysningene er anonymisert? *(Reidentifiseringstest, andre måter å fange opp om noe ikke er anonymisert?)*

12. Offentliggjør dere all statistikk? (*hvordan gjøres dette? Offentliggjør dere på deres nettside eller andre steder? Alt igjennom statistikkbanken? Er det statistikk dere velger å ikke offentliggjøre? Hvorfor ikke? Har dere innloggede tjenester? Hvordan blir disse autentisert? Hvordan sørger dere for at de som får tilgang til statistikk som ikke er offentliggjort ikke offentliggjør eller på annen måte misbruker tallene de får innsikt i? Er tilfeller der dere ikke kan gå god for at opplysningene dere offentliggjør eller på annen måte tilgjengeliggjør er anonymisert?*

13. Hvordan sørger dere for at opplysningene dere offentliggjør forblir anonyme? (*Herunder: økt bruk av stordata gjør at anonyme opplysninger dere offentliggjør kan sammenstilles med andre opplysninger som gjør at opplysningene ikke lengre er anonyme? Hvordan forholder dere dere til slike problemstillinger? Har SSB opplevd dette? Har dere en plan for om dette skjer?)*

14. Har dere vurdert om dere er pålagt å utføre en personvernkonsekvensutredning? (*Har dere utført en DPIA? (for eksempel ved*

en behandling som vil medføre en høy risiko for fysiske personers rettigheter og friheter?

15. Tar dere etter hensyn til *type* personopplysninger som statistikken baserer seg på? *(for eksempel personopplysninger av særlig kategori? andre sårbare grupper? Barn, eldre, psykiatri?)*