

Diskretion

Anvendelse af individdata
i Danmarks Statistiks forskerordninger

Forord

Danmarks Statistiks virke afhænger af, at der ikke i befolkningen og på virksomhederne opstår usikkerhed om, hvordan oplysninger afgivet til Danmarks Statistik bliver anvendt. Derfor lægger Danmarks Statistik stor vægt på at beskytte de mange fortrolige oplysninger om borgerne og virksomhederne. Dette fremgår bl.a. af diskretionspolitikken for både person- og erhvervsstatistikken.

Det er samtidig en højt prioriteret målsætning for Danmarks Statistik at data - uanset om de er følsomme og/eller detaljerede - skal kunne anvendes i forskningsprojekter. Derfor kan man som forsker få adgang til data under forudsætning af, at reglerne for datasikkerhed og diskretion efterleves.

Reglerne for datasikkerhed er nærmere beskrevet i *Datasikkerhedsreglement for Danmarks Statistik*, som er offentligt tilgængeligt på Danmarks Statistiks hjemmeside. Reglementet omfatter de generelle regler for, hvorledes oplysningerne skal beskyttes under opbevaring og bearbejdning i Danmarks Statistik.

Reglerne for diskretion sikrer, at det ikke er muligt at identificere hverken personer eller virksomheder i de statistikker, der publiceres af Danmarks Statistik. For forskerne har reglerne relevans fra det øjeblik oplysningerne printes ud eller hjemsendes, uanset om de skal publiceres eller blot er til eget brug. Reglerne for diskretion beskrives nærmere i det følgende.

Danmarks Statistik
Forskning og Metode
Sejrøgade 11
2100 København Ø
Telefon 39 17 39 17

Diskretionspolitikken

- Sværere at sikre anonymitet af virksomheder* Der er principielt ingen forskel på diskretionspolitikken for person- og erhvervsstatistikken. Men da der er meget færre virksomheder – og især meget få store virksomheder – end personer, er både direkte og indirekte identifikation af virksomheder lettere. Der er ofte kun en enkelt virksomhed inden for en given branche i et bestemt geografisk område, og problemet afhjælpes i sagens natur ikke af at bruge stikprøver.
- Indirekte identifikation* Datasæt, der er afidentificerede - hvor personnummer, virksomhedsnummer mv. er erstattet af et neutralt løbenummer – er ikke nødvendigvis anonymiseret. Dette skyldes at datasættene ofte indeholder så mange informationer, at det kan være muligt alligevel at finde frem til en person eller en virksomhed og slutte sig til oplysninger, som vedrører eller domineres af denne kendte person eller virksomhed.
- Det er vanskeligt at opstille enkle regler og det vil ofte være nødvendigt at vurdere, hvad der er rimeligt. Her er først beskrevet de generelle regler. Herefter følger et afsnit om personstatistik og sidst et om erhvervsstatistik.

Generelle regler

- Søg ikke efter personer* Det er ikke tilladt at lede efter enkeltpersoner (ikke en gang en selv) eller virksomheder – heller ikke som led i en vurdering af data.
- Tavshedspligt* Forskerne har tavshedspligt om oplysninger vedr. individer og virksomheder, som de måtte have fået igennem arbejdet med data – ligesom oplysningerne heller ikke må videregives. Reglerne om, hvorledes fortrolige data skal behandles, er dækket af Straffelovens § 152 og Forvaltningslovens § 27, stk. 3. Denne tavshedspligt vedvarer også efter forskningsprojektets ophør.
- Straffeloven §152* Stk. 1.: Den, som virker eller har virket i offentlig tjeneste eller hverv, og som uberettiget videregiver eller udnytter fortrolige oplysninger, hvortil den pågældende i den forbindelse har fået kendskab, straffes med bøde, hæfte eller fængsel indtil 6 måneder.
Stk. 2. Straffen kan stige til fængsel indtil 2 år, hvis forholdet er begået med forsæt til at skaffe sig eller andre uberettiget vinding eller der i øvrigt foreligger særlig skærpende omstændigheder.
Stk. 3. En oplysning er fortrolig, når den ved lov eller anden gyldig bestemmelse er betegnet som sådan, eller når det i øvrigt er nødvendigt at hemmeligholde den for at varetage væsentlige hensyn til offentlige eller private interesser.
- Forvaltningsloven - Tavshedspligt §27* Stk. 3.: En forvaltningsmyndighed kan bestemme, at en person uden for den offentlige forvaltning har tavshedspligt med hensyn til fortrolige oplysninger, som myndigheden videregiver til den pågældende uden at være forpligtet hertil.
Kilde: Karnovs Lovsamling 14. udgave, 1995, 1. & 3. bind
- Data må kun udprintes som statistik* Data må kun bruges til statistiske formål og det må ikke være muligt at identificere hverken personer eller virksomheder af statistikker, tabeller el.lign. Tommelfingerreglen er, at oplysninger, der printes ud eller hjemsendes skal være i publicerbar form, dvs. i form af tabeller, som principielt kunne tænkes offentliggjort uden reel risiko for identifikation af enkeltpersoner eller virksomheder.
- Se i Danmarks Statistiks publikationer* Det kan være en god ide at skele til Danmarks Statistiks publikationer, som kun indeholder tabeller, som ikke frembyder problemer. De er ikke særligt detaljerede og vil derfor ikke kunne give anledning til, at enkeltpersoner kan genkendes.

Personstatistikken

<i>Afhænger af følsomheden af data</i>	I personstatistikken er der i praksis ingen håndfaste regler. Det er et vurderings-spørgsmål, hvornår der skal diskretioneres. Der skal være mange personer i cellerne (dvs. der må være en grov eller aggregeret opdeling af data) hvis der er tale om følsomme data. Der findes ingen egentlig rangorden af Danmarks Statistiks person-variable, men de mest følsomme er oplysningerne om kriminalitet, anbringelser af børn, sygdomme (og andre sundhedsdata) samt sociale forhold.
<i>Data anvendt kommercielt</i>	En anden grund til diskretionering er af mere kommerciel art. Der er således en regel om, at der i geografiske inddelinger skal være mindst 1.000 beboere i hver tabelcelle, for at beskytte mod meget målrettet markedsføring.
<i>Let genkendelige karakteristika</i>	En tredje og mere simpel grund er blot, at det ikke skal være muligt at genfinde sig selv eller andre umiddelbart. Derfor skal data, der kan opdeles i mindre skarpt afgrænsede grupper (fx personer, med specielle uddannelser), diskretioneres for at garantere anonymiteten.

Erhvervsstatistikken

<i>To principper</i>	Diskretioneringspolitikken med hensyn til virksomhedsdata er mere præcis. Den udmønter sig i følgende to principper: Antalsprincippet og dominansprincippet.
<i>Antal enheder i cellerne</i>	Antalsprincippet angiver, at der ikke må offentliggøres informationer i en tabelcelle, der vedrører mindre end et bestemt antal virksomheder. Typisk skal der være mindst 3 enheder i tabelcellerne.
<i>Store virksomheder</i>	Dominansprincippet angiver, at der ikke må offentliggøres informationer i en tabelcelle, hvor de(n) største virksomhed(er) har en betydelig andel af tabelcellens samlede værdi. Denne regel sørger netop for at tage højde for den skæve størrelsesfordeling, som viser sig ved at et lille antal enheder tegner sig for en stor del af den samlede celleværdi. Typisk må to virksomheder ikke repræsentere mere end maksimalt 80 pct. af celleværdien.
<i>Grovere inddelinger eller udprækning af celler</i>	Diskretionering kan ske ved at indholdet i tabelceller udelades eller ved at lave grovere inddelinger eller grupperinger. Hvis diskretioneringen sker ved at "udprikke" tabelcellens indhold, skal man være opmærksom på, at cellen indirekte kan beregnes ved hjælp af totaler. Det er derfor som regel nødvendigt at udprikke flere celler.
<i>Pas på tabeller med opdeling på geografi og brancher</i>	Specielt bør man være yderst opmærksom, når der er tale om tabeller med opdeling på mindre geografiske områder, på enkelt-branche-niveau eller hvis der er tale om en kombination heraf – dvs. både geografisk opdeling og detaljerede brancheoplysninger.