

Forskningservice gennem et kvart århundrede

1. januar 1988 er en mærkedag i Danmarks Statistiks historie. Fra den dato kunne forskere fra danske forskningsinstitutioner få lov til at grave endnu dybere i data end tidligere: Danmarks Statistiks forskningservice var en realitet.

Siden da er antallet af tilknyttede forskere og forskningsmiljøer steget hurtigere, end nogen kunne forudse for mere end 25 år siden. Fra ganske få forskere i 1988 til mere end 1.400 i 2014. Derfor mener vi hos Danmarks Statistik, det er oplagt at fejre forskningservice. For uden denne åbne tilgang til forskerverdenen – naturligvis altid med datasikkerheden i top – ville meget forskning ikke kunne lade sig gøre, og meget viden om vores samfund og land ville ikke eksistere.

Denne udgivelse er ikke en kronologisk dag-til-dag-fortælling om, hvordan forskningservice har udviklet sig. I stedet laver udgivelsen nedslag i historien fra før etableringen af forskningservice og fortæller om de første, spæde år og om de begivenheder, der har formet forskningservice gennem det sidste kvarte århundrede

fortalt gennem de personer, som har haft forskningservice helt inde på livet.

Som det også vil fremgå af denne udgivelse, har en af de helt store personer omkring opbygningen og udviklingen af forskningservice været Otto Andersen, uden hvem forskningservice havde set noget anderledes ud i dag. Desværre gik Otto Andersen bort i januar 2014, hvor også hans mange anekdoter og store viden om forskningservice gik tabt.

Der skal lyde en stor tak til Elsebeth Lyng, Søren Leth-Sørensen, Knud Juel, Jan Plovsing, Peter Linde, Lars Thygesen, Henning Bunzel, Ivan Thaulow, Henrik Philip Jørgensen og Arne Brask Jørgensen, som med deres erindringer og gode historier har været helt uundværlige i arbejdet med at kigge tilbage på mere end 25 års historie.

God læselyst!

Indhold

Første forskningsskridt i 70'erne	4
Da startskuddet lød i 1988	6
Den næsten danske Nobelpris i økonomi	8
Kritik, gratis data og webadgang	10
Aarhus-filial til jyske forskere	14
Sociale forhold gav Københavns Kommune kvart milliard	15
Forskerbeskyttelsen og de faldende personstikprøver	17
Datasikkerhed er hjerteblod	18
Forskermaskinerne – teknikken bag	20
Lange tidsserier af høj kvalitet	22
Udlandet kigger med og kopierer den danske model	24

DANMARKS STATISTIK

28 NOV. 2014

BIBLIOTEKET

DA.25
709.8



Første forskningskridt i 70'erne

Danmarks Statistiks forskningsservice begyndte 1. januar 1988. Men faktisk strækker rødderne sig helt tilbage til 1970'erne. Her fik Elsebeth Lynges et stipendium af det daværende Statens Lægevidenskabelige Forskningsråd, og hun brugte sin tid på at sidde hos Danmarks Statistik for at undersøge sammenhængen mellem dødelighed og erhverv i forskellige socialgrupper.

- Vi kunne lave noget fantastisk ved at koble folketællingen fra 1970 og følge op med dødsfaldene senere hen, og det er klart, Danmarks Statistik var det eneste sted, man kunne lave sådan noget, siger Elsebeth Lynges, som i dag er professor på Center for Epidemiologi og Screening på Københavns Universitet. Hun var fra 1975-1978 forskningsstipendiat og sad i Danmarks Statistik. Kontakten opstod ved, at hun ringede til kontorchef i befolkningsstatistik, Poul Jensen, og spurgte, om det var muligt at sidde i Danmarks Statistik, og han var positiv overfor idéen.

- Det var på det her tidspunkt meget ultraditionelt, for der var en klar adskillelse mellem forskning og centraladministration. Det jeg lavede, var forskning, og Danmarks Statistik var centraladministration.

Jeg fik min løn fra Statens Lægevidenskabelige Forskningsråd, men sad hos Danmarks Statistik. Det var helt skævt dengang. Jeg kan ligeså tydeligt huske, der var en kontorfuldmægtig, som ikke mente, jeg kunne stå på telefonlisten, for jeg var jo ikke ansat. Det siger noget om, at det virkelig var starten på, at man gjorde det på den måde. Heldigvis var der nogen i Danmarks Statistik, som kunne se, at det var en god idé.

Det første større værk

Der er naturligt nok sket en omfattende positiv udvikling i kvaliteten af data fra 1970'erne til i dag. Fx var folketællingen fra 1970, som Elsebeth Lynges arbejdede med, aldrig blevet sorteret på nationalt niveau, men kun på amter. Dette skulle der først rådes bod på, og i den forbindelse viste det sig, at nogle borgere var registreret i flere amter. Først da der var ryddet op i dette, kunne borgerne kobles på CPR-registret.

- Så blev der fulgt op med dødsfaldene, og dødsfaldene fra 1970, 1971 og 1972 var ikke gemt med personnumre, så dem satte vi personnummer på efter en sundrig procedure. På den måde var det her det første større registerforskningsværk

nogensinde i Danmark, fortæller Elsebeth Lynges.

Omtrent samtidig med denne undersøgelse blev lavet i Danmark, iværksatte andre nordiske statistiske centralbureauer lignende undersøgelser. Efterfølgende overtog den på dette tidspunkt fremtidige kontorchef for den endnu ikke oprettede forskningsservice, Otto Andersen, opgaven med at lave opfølgningen.

ADAM skabte behovet

Der begyndte så småt at opstå forskellige projekter af lignende karakter. Der blev lavet flere efterfølgere til Elsebeth Lynges publikation, og der blev lavet en såkaldt fertilitetsdatabase, som eksisterer den dag i dag og indeholder henvisninger mellem børn og deres biologiske forældre. Og det var i disse tider grundidéen om en egentlig forskningsservice opstod, vurderer Elsebeth Lynges.

- I virkeligheden tror jeg egentlig, det var økonomerne, der skabte behovet for en forskningsservice i Danmarks Statistik, fordi de havde ADAM-modellen (en makroøkonomisk model ved navn Annual Danish Aggregate Model, red.).

Elsebeth Lyngemener, at en stor del af æren, for at Danmarks Statistik hurtigt udviklede sig til en forskerliberal institution, skal tildeles den nu tidligere kontorchef Otto Andersen.

- Det var en kæmpe fordel, at Otto Andersen havde en baggrund fra universitetsmiljøet. Han havde været lektor i demografi på Økonomisk Institut i mange år. Han blev glad for at være på Danmarks Statistik, og de blev glade for ham, og så havde han den forskningsbaggrund og dermed et lidt mere liberalt syn på forskere, end man traditionelt havde haft.

Ekstotisk projekt om renserier

Der er gennem årenes løb lavet mange banebrydende projekter i forbindelse med forskningsservice. Elsebeth Lyngem var selv med til at lave det, hun selv kalder et ekstotisk projekt, før forskerordningen blev etableret.

- Vi brugte folketællingen fra 1970, hvor vi kiggede på vaskeri- og renseriarbejdere. Men der er meget stor forskel på, om man arbejdede på vaskeri, som brugte vand, eller renseri, som brugte klorerede opløsningsmidler, siger Elsebeth Lyngem.

Resultaterne i en amerikansk undersøgelse tydede på, at sidstnævnte gruppe havde forøget risiko for nogle bestemte kræfttyper. Det ville Elsebeth Lyngem efterprøve i Danmark.

- I en udvalgt gruppe af kræftpatienter og nogle kontrolpatienter var vi faktisk nede og finde de oprindelige folketællingskemaer, hvor det står, om folk arbejdede i vaskeri eller renseri.

Sociale forskelle

Faktisk kunne forskerne i Danmark ikke eftergøre de amerikanske resultater, som især pegede på kræft i spiserøret hos de amerikanske renseriarbejdere. Og der er en god forklaring på forskellen mellem USA og Danmark.

- Renserier i Danmark var små, selvstændige virksomheder. Det var ikke en lavt betalt underklasse. Det var ligesom små købmænd, skotøjshandlere osv. Der var ikke nogen forøget risiko for spiserørskræft, og det passer godt med det, man ved om andre risikofaktorer for spiserørskræft, nemlig dårlig kost og højt alkoholforbrug. I USA havde man store renserier med meget lavt betalt, ufaglært arbejdskraft, så der var klart en forskel. Det var en helt anden gruppe, der arbejdede med kemikalier i Danmark. De var relativt godt uddannede og vidste en hel del om, hvad det var, de arbejdede med, siger Elsebeth Lyngem.

Udvikling uden sammenligning

Den dag i dag har Københavns Universitets Center for Epidemiologi og Screening flere projekter kørende på forskerordningen. De har netop afsluttet et projekt om livmoderhalskræft.

- Vi er i øjeblikket i gang med en anden undersøgelse, som går på, om kvinder, der får et abnormt svar i deres screening for livmoderhalskræft, har nogle langtidsvirkninger. Bruger de fx psykofarmaka, går de mere til læge, bliver de skilt, stopper de på deres uddannelse eller lignende. Det kræver jo kobling af en hel masse registre, og det kører helt uden problemer. De folk, der sidder i Danmarks Statistik, er jo også blevet dygtigere til det her, så jeg synes, det kører meget gelinde, lyder det fra Elsebeth Lyngem.

Udover at personerne omkring projekterne i årenes løb har dygtiggjort sig, har man i dag også en helt anden maskinkraft til at lave datakørsler sammenlignet med dengang.

- Hastighedsmæssigt er der overhovedet ingen sammenligning. Da man lavede folketællingen i 1970, kunne man ikke sortere hele befolkningen på én gang, kun amtsvist. I dag kan du sortere befolkningen på ingen tid. Det er så ændret, siger Elsebeth Lyngem.

- Så må man også sige, at Danmarks Statistik har været gode til at følge med og lægge alt muligt software, som forskerne har brug for, til rådighed. Derudover er der blevet bedre datadokumentation. Man har brug for god dokumentation på data, specielt over tid, for der kan jo have været brud i dataindsamling eller på definitioner.



Da startskuddet lød i 1988

1. januar 1988 havde Søren Leth-Sørensen første arbejdsdag i Danmarks Statistiks forskningservice. Han var hyret til at sætte projektet i gang i samarbejde med Ruth Emerek, Per Vejrup-Hansen og Otto Andersen. Disse fire blev sammen med Finn Spieker flyttet ud af huset og til andres lokaler i Kristineberg på den anden side af gaden, hvor man lavede et forskningsmiljø med lovmodellen og forskellige forskere fra bl.a. Rockwool Fonden.

- Idéen var, at nu sætter vi sådan nogle langhårede nogle sammen der, og så håber vi, at vi får noget ud af det. Personligt fik jeg meget ud af, hvordan lovmodellen dokumenterede deres data. Det var også her, jeg blev præsenteret for den første pc. Det var jo et værre vidunder, siger Søren Leth-Sørensen.

Sensationen tre år tidligere

Tre år tidligere – i 1985 - var der nordisk statistikermøde i Hanstholm om arbejdsmarkedsstatistik. En af deltagerne på mødet var Sven Egmoser, som i dag er kontorchef i Danmarks Statistik. Søren Leth-Sørensen deltog også som ansvarlig for den registerbaserede arbejdsstyrkestatistik (RAS).

- Sven havde lavet sådan noget, hvor man kunne følge personers beskæftigelse over tid. Fx kunne man se, hvem der det ene år var i beskæftigelse og året efter var arbejdsløs og omvendt. Det var noget nyt i 1985, siger Søren Leth-Sørensen.

Med på mødet var også Per Vejrup-Hansen, som på daværende tidspunkt var ansat på SFI. Her fortalte han om en undersøgelse, hvor han havde undersøgt jobmobiliteten på arbejdsmarkedet fra et år til det næste.

Man bruger oplysningerne fra SKAT, hvor der foreligger en oplysning om de firmaer, hvor en person har været ansat i løbet af året.

- Men virksomheder har det med at skifte registreringsnummer, så hvis man blot bruger dette til at følge virksomhederne med, så kan det se ud som om en virksomhed forsvinder fra et år til det næste, og samtidig ser det ud som om, en virksomhed er dukket op. Så bliver der en masse hurlumhej i statistikken med folk, der tilsyneladende skifter job, uden at det reelt er tilfældet, siger Søren Leth-Sørensen.

Styr på hurlumhej

For at råde bod på dette statistiske hurlumhej brugte IDA-gruppen forskellige indikatorer for at se, om der var tale om en ny virksomhed eller blot en gammel med nyt registreringsnummer. Hvis man har oplysninger om virksomheder på to på hinanden følgende år, og der er et sammenfald geografisk og branchemæssigt, så kunne det være et tegn på, at det var den samme virksomhed. Per Vejrup-Hansen havde også brugt personsammenfald som en yderligere indikator på, at der var tale om en og samme virksomhed. For en given virksomhed er der ofte et personsammenfald på 70-80 pct. fra et år til det næste. Hvis der var en stor andel gengangere i to virksomheder med forskellige registreringsnumre, så kan det være yderligere en indikator på, at der var tale om den samme virksomhed.

Der opstod hurtigt enighed om, at Danmarks Statistik burde prøve at danne data, der var bedre til at analysere jobmobilitet og stille dem til rådighed for forskerne. Danmarks Statistik valgte at spørge alle de betydende forskningsinstitutioner, om et samarbejde om et sådan projekt ikke var en god idé, og de sagde ikke overraskende alle ja.

- Så lå det implicit, at det skulle være noget, forskerne kunne bruge.

Dermed var finansieringen på plads, og Otto Andersen blev kørt i stilling som chef.

Brugsen på Sallingvej

Som det er med nye projekter, så kan der opstå uenighed. Det gjorde der også, da forskningsservice var gået i luften.

Konflikten i dette tilfælde gik i al sin enkelthed ud på, hvordan man definerer, om en virksomhed er en ny virksomhed, eller om det er gammel, der blot har skiftet registreringsnummer.

- Vi brugte Brugsen på Sallingvej som eksempel. Det ene synspunkt var, at der kun kunne være én definition på, hvornår en virksomhed var den samme, mens det andet synspunkt var, at det afhænger af den problemstilling, man er interesseret i, siger Søren Leth-Sørensen.

- Man kan se på det juridisk: Hvis forgængereren ikke betalte lønmodtagerens skat ind til skattevæsenet, kan man så gøre hans efterfølger ansvarlig? Man kan se det fra borgerens synspunkt: Han er ligeglad med, om butikken hedder Brugsen, Spar eller noget tredje. Der skal bare være en købmandsforretning. Så der kan anlægges flere definitioner på, hvornår noget nyt er nyt i denne sammenhæng.

IDA bliver født

Forskningsservices første barn bliver døbt IDA - den Integrerede Database for Arbejdsmarkedsforskning. Formålet med IDA var, at man skulle kunne følge virksomheder over tid, ligesom man kan med personer.

- Så opgaven var at finde en model, der kunne følge virksomhederne, så vi kunne se, om de eksisterede fra det ene år til det andet. Og det kan man altså ikke se, når man sidder og kigger på nogle numre. Derfor fandt vi ud af, vi kunne kigge på personsammenfald, branchesammenfald og geografisk beliggenhed.

Efter tre års arbejde var IDA færdig, og 1. januar 1991 var forskningsservice klar til at tilbyde data til forskerne.

På mission gennem Danmark

I løbet af de tre år stod arbejdet dog ikke kun på programmering af databasen. Selvom vi er mere end 20 år tilbage i tiden, så levede forskningsservice allerede dengang op til nuværende rigsstatistiker Jørgen Elmeskovs krav om, at Danmarks Statistik skal være en åben institution overfor omverdenen.

- En af de specielle ting var, at mens vi lavede IDA-databasen, tog vi rundt om i landet, bl.a. til Aarhus, Aalborg og Odense, for at fortælle om, hvad vi var i gang med at lave, men også for at høre hvilke ønsker brugerne havde. Der var jo nogle folk i Danmarks Statistik, der i 1988

mente, at vi kun skulle lave det, der stod i Lov om Danmarks Statistik. Og kunder! Det skulle vi ikke beskæftige os med!

PC'ere gjorde livet lettere

Specielt var det, at Danmarks Statistiks forskningsservice nu besad tre pc'ere. En til deling mellem Ruth Emerek og Per Vejrup-Hansen, en til sekretæren Bente Biering-Sørensen og en til Søren Leth-Sørensen. Og pc'ernes indtog gjorde livet meget lettere.

- Ellers skulle vi jo sidde og skrive på en skrivemaskine. Så har man et stykke papir, som du kan fotokopiere, men du kan ikke lave nogen systemer af en årt, sortere eller redigere i en tekst. Der måtte man sidde og klippe-klistre. Det første, vi tænkte, var, det var da smart, man kunne redigere. Men selvfølgelig også det med, at man kunne få en diskette. Så kunne man, hvis man havde en rapport, have den med sig. Det var en kæmpe fordel.



Den næsten danske Nobelpris i økonomi

Det var altså i slutningen af 80'erne, at arbejdet med IDA – den Integrerede Database for Arbejdsmarkedsforskning – blev påbegyndt. Og mere end to årtier senere skulle IDA få sin plads på den største, økonomiske sejrskammel af dem alle. Æren herfor skal tilskrives den amerikanske økonom, Dale T. Mortensen – hvis efternavn skyldes en dansk far – som i en årrække var tilknyttet Aarhus Universitet. Han blev tiltrukket af Danmark på grund af de store muligheder inden for registerforskning, og i 2010 modtog han Nobelprisen i økonomi, blandt andet på grund af sit arbejde med IDA.

Begyndelsen 27 år tidligere

- Det begyndte egentlig helt tilbage i 1983. Dengang havde Danmark som et af de første lande fået individoplysninger på ledighed helt ned på ugebasis. Man var begyndt at tale om, at der ikke var en fast bestand af arbejdsløse, men der var mange strømninger ind og ud af job, fortæller Henning Bunzel, som er lektor på Aarhus Universitet og var tilknyttet Dale T. Mortensens projekt.

Sådan var udgangspunktet, og næsten 20 år efter havde Dale T. Mortensen og hans team udviklet modeller til at beskrive

sammenhængen mellem virksomheders vækst, arbejdsløshed, ledige jobs og lønninger, så man blandt andet kan finde ud af, hvorfor der er ledige stillinger, hvis der samtidig er stor arbejdsløshed. I 2010 blev modellerne fundet værdige til Nobelprisen i økonomi. Det skyldtes hans teoretiske arbejde, som skete i 1980'erne og 1990'erne, men at disse modeller blev udbygget, så de passede til danske data, var stærkt medvirkende til den flotte pris.

- IDA var jo en meget værdifuld måde at få organiseret oplysninger om arbejde og arbejdssteder, så det har virkelig været kernen i data helt frem til begyndelsen af 2005, siger Henning Bunzel.

Helt nyt på verdensplan

IDA var speciel, da den blev lanceret, for det var første gang, man på verdensplan havde en database, som koblede et lands totalpopulation og virksomheder.

- Det, IDA gjorde så godt, var at få så meget ud af bruge de registerdata, som var samlet ind til forskellige administrative formål og så bearbejde dem til at have det økonomiske indhold, som vi havde brug for. Der er mange variable, som man konstruerede på baggrund af registrene, som ikke var noget, man direkte havde obser-

veret, fx timeløn, fortæller Henning Bunzel.

En amerikaner valgte Danmark

Det var ikke en tilfældighed, der gjorde, at Dale T. Mortensen valgte, at han skulle forske i Danmark og ikke i hjemlandet USA. Tilgængeligheden af data var altafgørende for amerikaneren, der var professor ved Northwestern University, men hvis andet hjem i en årrække var Institut for Økonomi på Aarhus Universitet.

- Problemet i USA var, at mange af de virksomhedsdata kommer fra enkelte stater, så der skal man ud i den enkelte stat og have tilladelse til at bruge dem. Så processen var meget tung, og jeg tror egentlig aldrig Dale og hans medforfatter fik adgang til de amerikanske data, hvilket han egentlig havde håbet på. Amerikanerne har haft nogle meget gode stikprøver, men det er jo stadig en meget lille population, når man arbejder med en model, hvor man kan se, hvordan folk flytter rundt i beskæftigelse, og hvordan virksomhederne udvikler sig over tiden. Det er derfor helt umuligt at lave en fornuftig stikprøve, fordi der er tale om, at tingene bevæger sig i to dimensioner. Derfor er det meget væsentligt, at vi har adgang til totalpopulationen siger Henning Bunzel.

Dermed er der stor forskel på måden, man har valgt at køre forskningsservice på i de amerikanske statistikkontorer sammenlignet med Danmarks Statistik.

- Det danske system er fleksibelt, fordi man har accepteret, at folk sidder og arbejder på data i deres eget kontor gennem sikre forbindelser. Det gør det altså meget nemmere i hverdagen, end hvis man skulle flytte sig et eller andet sted hen for at sidde og arbejde på data, som man hidtil har skullet i andre lande, siger Henning Bunzel.





Kritik, gratis data og webadgang

Man skulle tro, at forskerne rundt omkring på forskellige institutioner klapper i hænderne af begejstring over Danmarks Statistiks forskningsservice. Men sådan har det langtfra altid været.

Jan Plovsing, som var rigsstatistiker fra 1995 til 2013, husker, hvordan forskerne tilbage i januar 2000 kritiserede Danmarks Statistik ved Registerforskningskonferencen, som fandt sted i Landstingssalen. Blandt deltagerne var daværende forskningsminister Birthe Weiss, en lang række forskere og Jan Plovsing. Konferencens forhistorie var, at der havde været kritik fra forskerne i pressen over, at man skulle være fysisk til stede i Danmarks Statistiks lokaler for at få adgang til at analysere data.

Tre kritikpunkter fra forskerne

- Kravet om den fysiske tilstedeværelse var den ene ting, forskerne var utilfredse med. Og så var de utilfredse med priserne på data og utilfredse med priserne for at køre på forskermaskinen. Det var sådan set de tre kritikpunkter af os, siger daværende rigsstatistiker Jan Plovsing, som ene mand måtte forsvare Danmarks Statistik den sene januar-dag i Landstingssalen.

- Der var 8-9 indlæg fra forskellige forskere, og så var der en til at forsvare os, det var mig, og jeg blev kørt godt igennem, siger Jan Plovsing.

Her blev det også luftet, hvordan Sundhedsstyrelsen på daværende tidspunkt leverede data til forskerne. Den var nemlig anderledes end praksis hos Danmarks Statistik.

- De udleverede simpelthen data med identifikationer til forskerne, og det var forskerne meget tilfredse med. Men de var utilfredse af andre grunde, for de fik nemlig oldgamle data, og kvaliteten var lav og dokumentationen dårlig. Vi leverede allerede dengang et høj kvalitetsprodukt, men ville have penge for det. Sundhedsstyrelsen leverede et produkt, som var gratis, men af lavere kvalitet.

Forskere ville have CPR

Det har aldrig været på tale for Danmarks Statistik at udlevere data med personoplysninger. Sikkerhed af borgernes og virksomheders data vægtes over forskernes tilfredshed.

- Det var i virkeligheden det, det hele gik ud på, lyder vurderingen fra Jan Plovsing.

- De ville have data udleveret med CPR-nummer eller anden identifikation, som de fik det hos Sundhedsstyrelsen, så de kunne koble det på deres eget system. Men de formulerede det på en anden måde: De forskermaskiner, som de analyserede på, skulle have deres egen bestyrelse, som ene og alene bestod af forskere. Og så ville de have en direkte opkobling til Danmarks Statistik, så de ikke behøvede at bevæge sig herhen, og så skulle alt være meget billigt.

Indtægtsdækket virksomhed

At Jan Plovsing ville komme i krydsild denne mandag, dagen før januar blev til februar, var han godt klar over, inden han mødte op. Det vigtigste budskab til forskerne var, at forskningsservice hørte ind under statens regler for indtægtsdækket virksomhed, hvilket betyder, det ikke må give overskud, så faktisk tjente Danmarks Statistik ikke penge på datakørslerne. Men det må heller ikke give underskud, for så bruger Danmarks Statistik sin bevilling i strid med loven.

- Mange af indlæggene gik på, jamen I har jo data i forvejen, vi må kunne få nogle marginalpriser. Så måtte jeg jo forklare, at alle omkostninger skal være dækket –

både de direkte og de indirekte. De mente jo højst, at det var de direkte.

Det kunne der altså ikke røres ved. Det var lovbestemt, at regnskabets slutresultat skulle gå i nul, når året var omme. Men hvis forskerne havde forventet at opleve en rigsstatistiker, som kun parrerede de slag, der kom ind, så blev de skuffede denne mandag. Positivt skuffede. For Jan Plovsing benyttede lejligheden til at gå i offensiven.

Forskningservice ér dyrt

- Jeg sagde, det er da rigtigt venner, det er dyrt. For vi har 12 ansatte til at levere service til jer. Og hvad kan vi gøre ved det? Der må jo komme nogen med nogle penge udefra. Det kan være fra Folketinget, Regeringen eller Grundforskningsfonden.

Hvis disse penge blev bevilliget, kunne det altså gøres gratis for forskerne. Mange data var allerede organiseret i forskningsdatabaser, fx IDA, og hvis nogen ville betale for at ajourføre disse, så ville forskerne naturligt nok kunne spare penge.

I forbindelse med dette forslag sparkede Jan Plovsing også en anden dør åben, nemlig døren til Statistikbanken. På dette tidspunkt var den kun tilgængelig for betalende brugere. Med argumentet om, at det var demokratisk ikke så holdbart, men at Danmarks Statistik ikke havde råd til at gøre den gratis tilgængelig, blev også Statistikbanken altså sendt i spil. Så den kunne blive tilgængelig for alle borgere og

ikke kun for dem, der kunne betale.

- Budskabet var i virkeligheden fra min side, at vi gerne imødekommer ønskerne – dog ikke det med udlevering af identificerbare data, og jeg nævnte heller ikke direkte adgang, da vores forståelse af sagerne var, at det ville være for usikkert.

Jan Plovsing havde seks dage i forvejen været til møde med daværende økonomiminister Marianne Jelved og forskningsminister Birthe Weiss for at tage forskud på, hvad der ville ske og se, om der var mulighed for et kompromis, så det ikke blev alt for konfliktfyldt.

Fra forslag til forhandling

Der var ikke langt fra tanke til handling hos Danmarks Statistik eller de involverede ministerier. Allerede i foråret 2000 indledte Jan Plovsing forhandlinger med Finansministeriet og Økonomiministeriet, og da sommeren kom, var en aftale om en finanslovsbevilling på plads. Den ville gøre Statistikbanken, *Nyt fra Danmarks Statistik* og webudgaven af *Statistisk Årbog* gratis tilgængelig.

- Det lykkedes os at få en bevilling på 9 mio. kr. årligt til dette, og det betragter jeg som en ret god handel for os og samtidig for befolkningen i Danmark. Både for menneskene og pressen var en helt masse ting pludselig tilgængelig, og det var en stor succes.

Ekstern mailadgang

2000 var i det hele taget et år, hvor der skete en helt masse i Danmarks Statistik. I juni måned skete der noget, som senere skulle få stor betydning for forskningsservice. Denne juni udsendte It-center en meddelelse om, at medarbejderne fremover med det netop implementerede mailsystem kunne tilgå deres e-mailadresse via internettet hjemmefra, eller når de var på rejse.

- Det var sikkert nok, når man bare sikrede forbindelsen og brugte den token, vi havde på det tidspunkt. Det var sådan en nøgle med nogle tal, der ruller, og hvert minut kom et nyt tal op. Så skulle man bare indtaste kode og tal, fortæller Jan Plovsing.

Dermed var Danmarks Statistiks medarbejdere pludselig langt mere mobile, da de kunne tilgå både mail og kalender eksternt. Og denne mulighed skulle også snart smitte af på forskningsservice.

Opringing fra Ålandsøerne

Da juni var blevet til juli, som var blevet til august, var det tid til det årlige, nordiske chefstatistikermøde, som i 2000 blev afholdt på Ålandsøerne. Her havde Jan Plovsing tid til at tænke over denne nye onlineadgang til e-mail og kalender.

- Jeg har måske haft tid til at sidde og filosofere mellem indlæggene. I hvert fald fik jeg den idé, at når det nu var sikkert nok at få adgang til vores mail via det system, så måtte det også være sikkert nok

til, at forskerne kunne få adgang til data fra deres arbejdsplads med et tilsvarende system. Jeg kunne ikke helt se forskellen, siger Jan Plovsing 14 år efter dette chefstatistikermøde.

Så han greb telefonen og ringede til Otto Andersen. Og det var en meget målbevidst og bestemt rigsstatistiker, som Otto Andersen fik i tale.

- Så sagde jeg til Otto: Nu skal du høre. Jeg ved, der vil være nogen modstand her i huset mod det her, og alle, der beskæftiger sig med sikkerhed og data, vil ikke være begejstrede, men ligeså snart jeg kommer hjem, så har vi nedsat et udvalg, og du kan godt begynde at tænke på nogle personer. Og kommissoriet skal ikke hedde, om forskerne skal have adgang. Jeg gider ikke diskutere, om de skal have adgang, den beslutning har jeg truffet, men hvordan de skal have adgang, tænker Jan Plovsing tilbage.

En sådan kommando fra en overordnet bifaldes ikke nødvendigvis altid af modtageren, men Otto Andersen var i forvejen tilhænger af webadgang til forskerne. Så allerede i løbet af efteråret var det udredt, hvordan den tekniske løsning skulle være.

Autorisation er vejen ind

Ligesom i dag var det dengang udelukkende autoriserede forskningsmiljøer og autoriserede forskere, som fik adgang til Danmarks Statistiks forskningsdata.

For at få autorisation kræves der først og fremmest, at Danmarks Statistik vurderer, at et forskningsmiljø er tilstrækkeligt stort og anerkendt. Når der er sat flueben ud for dette, skal den ansvarlige chef på stedet skrive under på en erklæring om vilkårene for behandling af data, samt at forskerne på stedet ikke søger efter enkeltpersoner. Denne erklæring underskrives også af hver enkelt forsker.

- Der er så indbygget nogle ret stramme sanktioner, hvis man bryder dette. Hvis nu en forsker fra en institution bryder reglerne, så kan vi udelukke alle forskere fra forskermiljøet i flere år fra overhovedet at få adgang til vores data, siger Jan Plovsing.

Forskere får deres vilje

Det arbejde, som for alvor blev påbegyndt i efteråret 2000, var fuldent i starten af 2001, og efter godkendelse fra Danmarks Statistiks styrelse kunne forskerne fra 1. marts 2001 få adgang til at analysere data direkte fra deres arbejdsplads. Dog med nogle få undtagelser.

For i første omgang var virksomhedsdata udeladt, ligesom der var to undtagelser på det sociale område, nemlig anbragte børn og unge samt kriminalitetsstatistikken. Det skete af sikkerhedshensyn.

- Man betragtede i Danmarks Statistik anbragte børn og unge samt kriminalitetsstatistikken som særligt fortrolige, og med hensyn til virksomhedsstatistikken så er det lettere at identificere virksomheder,

end det er med personer. Så det gjorde vi til at starte med, og så måtte vi samle nogle erfaringer. Det fungerede jo, og så blev det udvidet et par år efter, siger Jan Plovsing om den beslutning, som styrelsen traf i 2003, hvor forskerne også fik webadgang til disse statistikker, dog med få undtagelser for virksomhedsdata.

6 mio. kr. fjerner sidste knast

Med foråret 2001 godt på vej kunne forskerne glæde sig over, at de nu kunne tilgå forskningsdata fra deres arbejds-pc og var forskånet for en cykel- eller køretur til Danmarks Statistik for at arbejde med data. Tilfredsheden havde dog ikke toppet – for det var stadig alt for dyrt. Men samtidig med, at arbejdet med webtilgængelighed var startet, havde man også nedsat et registerforskningsudvalg efter aftale med forskningsminister Birthe Weiss. Udvalget bestod af en række centrale forskere, Otto Andersen og Jan Plovsing og skulle undersøge, hvordan man gjorde det billigere og måske endda gratis for forskere at arbejde med data fra Danmarks Statistik. Udvalget startede sit arbejde i november 2000 og var færdig i juni 2001.

- Nu havde forskerne fået den ene af de ting, de ønskede, men de andre ting omkring hvor dyrt det var, det var sådan set ikke klaret. Otto og jeg skrev en masse papirer for at finde ud af, hvad det ville koste, og da udvalget sluttede sit arbejde i juni 2001, kom der en finanslovsbevilgning fra Forskningsministeriet på omkring 6 mio. kr. årligt, siger Jan Plovsing. Dermed

havde forskerne endelig fået opfyldt alle deres ønsker eller krav, om man vil.

- For de penge kunne finansiere det, de havde krævet dengang: Gratis kørsler på forskermaskinen, opdaterede forskningsdatabaser, gratis rådgivning inden for rimelige grænser osv. Så tilsammen har det gjort, at vi har fået nogle penge at arbejde med, så vi kan servicere forskerne meget bedre.

Bevillingen kom første gang 1. januar 2002, og her 12 år efter eksisterer den stadig. Den administreres og kontrolleres af det Koordinerende Organ for Registerforskning (KOR).

- Vi har faktisk ikke hørt nogen nærværdig kritik siden, men selvfølgelig er forskere per definition aldrig tilfredse. Det ligger i forskningens natur, at man skal være lidt kritisk, lyder den latterfulde vurdering fra Jan Plovsing.

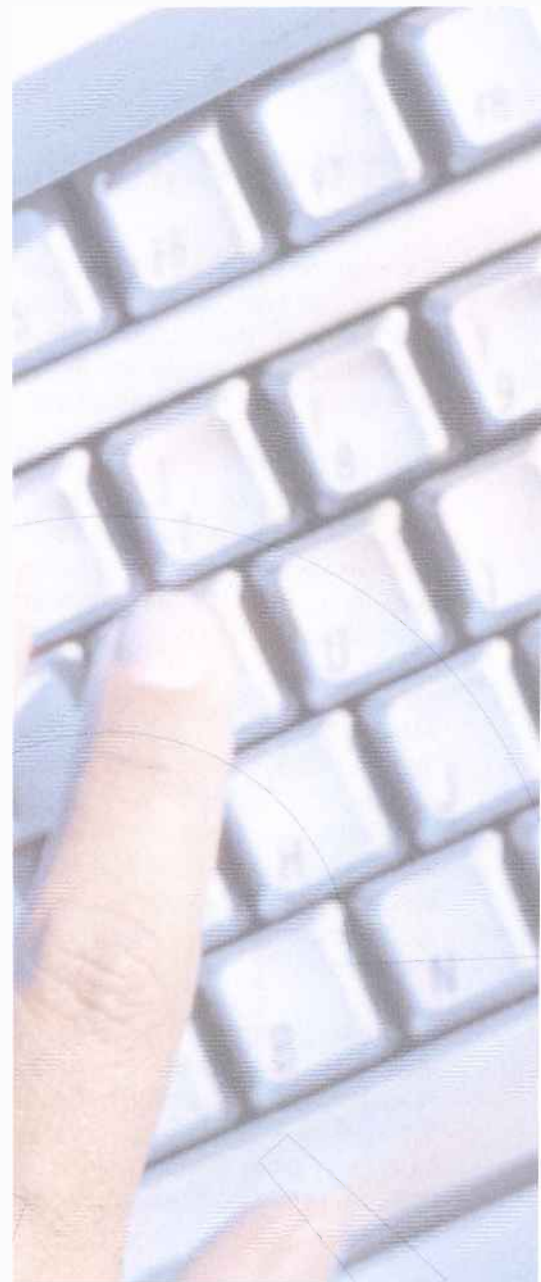
- Siden er ordningen også blevet udvidet. Fx kan udenlandske forskere, der er tilknyttet en dansk, autoriseret institution også få lov til at bruge forskningsservice.

Godt for samfundet

Mængdemæssigt har brugen af Danmarks Statistiks forskningsdata været konstante stigende fra dag et. Det skyldes, at der er et behov og mange tilfredse kunder i butikken, mener Jan Plovsing. Brugen af denne service breder sig på positiv vis, som de berømte ringe i det våde vand.

- Det er godt for samfundet, for meget af forskningen har jo en umiddelbar nytte-værdi, fx på sundhedsområdet. Jeg kan huske, at der tilbage i tiden var kritik af Lægemedelstyrelsen, som jo har en kæmpe database med folks medicinforbrug. Det var jo indlysende, at kombinerer det med en række af vores sociale data, så får man nogle effektstudier, som både lægevidenskaben og som industrien var interesseret i, men de ville ikke udlevere data eller etablere noget som helst til forskerne, siger Jan Plovsing om en af de kampe, forskningsservice har måttet tage gennem sin levetid.

- Men i begyndelsen af 00'erne blev vi enige med Lægemedelstyrelsen, så disse data kunne bruges af Danmarks Statistik. Lægemedelstyrelsen skulle give nogle tilladelser, foruden dem Danmarks Statistik skulle give til forskerne. Derefter kunne Lægemedelstyrelsens data komme til at indgå som en database, vi også kunne benytte os af. Det var en cadeau til Danmarks Statistiks forskningsservice. Så vi har på mange måder fået mere og mere at gøre godt med datamæssigt og udvidet brugen af det til forskerne.





Aarhus-filial til jyske forskere

De fleste forbinder Danmarks Statistik med den massive, røde bygning på Østerbro, København. Men faktisk havde Danmarks Statistiks forskningsservice fra 1996 til 2014 også en filial på Aarhus Universitet.

- Den opstod, fordi vi af sikkerhedsmæssige grunde krævede, at man fysisk skulle sidde i Danmarks Statistik for at arbejde med data. Der var langt over til København for jyderne, fortæller daværende rigsstatistiker Jan Plovsing.

Fem-årig bevilling

Aarhus-filialen blev i starten finansieret med en fem-årig bevilling fra Grundforskningsfonden. I 2001 udløb denne, og så skulle betaler vurdere, om han fik nok for sine penge. Derfor nedsatte de et udvalg bestående af to uafhængige eksperter. I 2002 tikkede evalueringen ind, og den var yderst positiv.

Problemet i gåseøjne var dog på dette tidspunkt, at Danmarks Statistik året forinden havde fået tildelt en finanslovsbevilling på omkring 6 mio. kr. af Forskningsministeriet for at levere gratis data og rådgivning til forskere, og derfor stoppede bevillingen.

Levede videre uden bevilling
Men lukkes skulle filialen ikke, mente man på Aarhus Universitet, og man nåede på let vis til enighed om en praksis, som ikke ville koste Danmarks Statistik mere, end hvis de havde haft de to daværende Aarhus-medarbejdere siddende i København.

- Enden på det hele blev, at filialen kørte videre, men at Aarhus Universitet stillede lokaler gratis til rådighed, hvorimod vi betalte lønnen, fortæller Jan Plovsing.

Overlevede onlineadgang

1. marts 2001 lukkede Danmarks Statistik op for forskernes onlineadgang til data, men faktisk lukkede man først Aarhus-filialen 13 år efter. I løbet af disse mange år har Søren Leth-Sørensen set en stor teknisk udvikling i, hvordan man kan servicere forskerne.

- De her dataudtræk er blevet noget nemmere at foretage. Vi har investeret ret mange timer, knofedt og vilje for at gøre det meget nemmere at lave udtræk til forskerne. Det er jo alt sammen skræddersyet. I 1996 brugte man de programmer, der var udviklet i Danmarks Statistik. Det var med båndkørsler, og det tog en evighed. Dengang havde vi heller ikke data samlet, så

der måtte vi henvende os til fagkontorerne. Det er blevet 100 gange nemmere, siger Søren Leth-Sørensen.

Det blev altså besluttet i 2014, at Aarhus-filialen skulle lukkes, da forskerne alligevel sidder på deres arbejdspladser i dag og ikke som tidligere møder fysisk op. Denne lukning har betydet en ny, spændende udfordring til Søren Leth-Sørensen, som i stedet for at være udstationeret i Aarhus, bliver udstationeret i efteråret 2014 som rådgiver i Bosnien-Hercegovina i et stort projekt til forbedring af statistikken.



Sociale forhold gav Københavns Kommune kvart milliard

Danmarks Statistiks forskningsservice har i de senere år oplevet en enorm stigning i antal autoriserede forskere, men den første forsker, som benyttede sig af Danmarks Statistiks forskningsservice, var Knud Juel. I dag er han forskningsleder på Statens Institut for Folkesundhed, men for 25 år siden var han cand.stat. hos Dansk Institut for Klinisk Epidemiologi. Og dengang fungerede det væsentligt anderledes sammenlignet med i dag, hvor forskerne har onlineadgang til data fra deres institution.

- Jeg havde adgangskort, nøgle og et kontor, men oftest sad jeg der efter kl. 22 om aftenen, for der kørte Danmarks Statistik med nogle lavere priser. Og det skulle foregå på Danmarks Statistik, fordi alle data lå der, siger Knud Juel her godt 25 år efter.

Her sad han og lavede datakørsler, som blevet grundlaget for publikationen *Sygehusindlæggelser og sociale forhold*, som i 1990 kortlagde befolkningens indlæggelsesmønster på sygehuse.

- De statistiske analyser handlede om at se på, hvorfor der var en overhyppighed i indlæggelser i København og Frederiksberg Kommuner sammenlignet med de

andre daværende amter. Jeg brugte landspatientregistret og Danmarks Statistiks registre og kørte på individniveau. Vi tog så hensyn til de individuelle karakteristika køn, alder, civilstatus, uddannelse osv. Dermed havde vi oplysninger om hver enkelt dansker, og så kiggede vi på indlæggelser og sygedage, siger Knud Juel.

Konflikt til kvart milliard

- Vi ville se, om personer i København og på Frederiksberg var mere syge eller mere pivede. Hvis der fx var flere enlige mænd i København eller flere personer med mindre eller ingen uddannelse, som kan forklare overforbruget, jamen så var de ikke pivede, så var de syge. Så var København og Frederiksberg bare uheldige med deres befolkningssammensætning, og det skulle de ikke bøde for.

Konflikten var kort sagt, at Københavns Kommune mente, de fik for få penge til sygehusvæsenet af staten, som på den anden side mente, at Københavns Kommune fik nøjagtigt et passende beløb.

- Denne opgave kunne bedst løses ved at lave denne slags analyse, for ellers kunne Københavns Kommune blive ved med at hævde, at de skulle have flere penge på

grund af befolkningssammensætningen, og Amdrårdsforeningen ville bare sige, jamen I sløser med midlerne.

Resultatet blev, at statistikken talte sit tydelige sprog: Københavnerne var ikke pivede, de var mere syge. Dermed fik Københavns og Frederiksberg Kommuner tilsammen overført en kvart milliard kr. i ekstrabevilling fra staten.

Umuligt uden statistikken

Der er altså en stor sammenhæng mellem forskningen og Danmarks Statistik, for ofte er sidstnævnte nødvendig for at førstnævnte kan levere sine resultater.

- Man kunne ikke have lavet vores publikation uden Danmarks Statistik. Vi havde brug for de karakteristika (køn, alder, civilstatus, uddannelse osv., red.), og de oplysninger findes kun hos Danmarks Statistik, siger Knud Juel.

- Det handler om, at hvis man lader som om, der er lige mange 40-årige, lige mange mænd, lige mange kortuddannede, lige mange ugifte, er der så forskel på København og resten af landet. Man laver den direkte sammenligning. Man kunne være ret sikker på, før vi lavede analysen, at for-

skellen ville blive mindre, for vi kunne jo godt se i statistikkerne, at der var relativt flere kortuddannede i Københavns Kommune, men vi vidste ikke, hvor meget det ville betyde, når vi tog det hele på en gang. Det kunne være lidt, og det kunne være det hele.

Det samme og alligevel ikke

Der er gennem årenes løb sket en del forbedringer i forskningsservice, lyder dommen fra Knud Juel, selvom han som forskningsleder ikke længere selv bruger ordningen så meget som tidligere. Det har han overladt til sine kolleger.

- Man har terminaladgang fra sit eget skrivebord, det er selvfølgelig en stor fordel, og så er maskinerne blevet større. Der kan også laves analyser i dag, som tager lang tid, men det er ikke noget, der minder om dengang.

Det var også derfor, Knud Juel måtte bruge sene aften- og nattetimer i slutningen af 1980'erne på at sidde og lave kørsler i Danmarks Statistik.

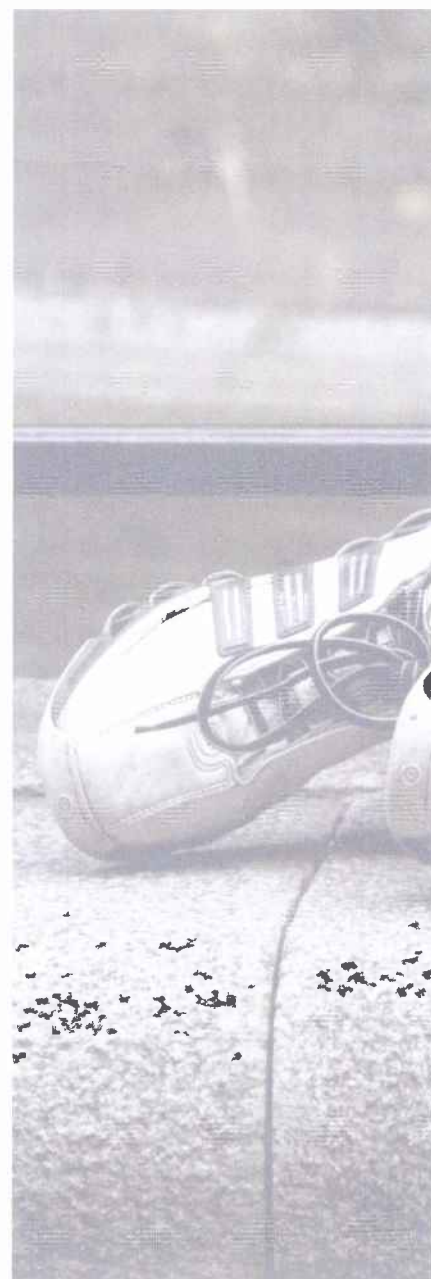
- Det var store statistiske analyser på store datamængder. Nu om dage ville man ikke sige, det var noget stort. Der ville man kunne lave det på en pc eller en bærbar, men det var maskinkraften, der skulle udnyttes, og den blev frigivet om natten.

Særlige muligheder i Danmark

For Statens Institut for Folkesundhed og mange andre forskningsinstitutioner er Danmarks Statistiks forskningsservice i dag ligesom for 25 år siden en unik indgang til at lave undersøgelser, som ellers havde været umulige at gennemføre.

- Vi kan lave analyser, som simpelthen ikke kunne lade sig gøre ellers. Vi har jo fokus på sundhedsdata, og dem ville vi kunne lave mange spændende ting med, men ved at koble sundhedsdata med Danmarks Statistiks data får vi den sociale dimension med ind i det. Vi kan fx se, hvor meget uddannelse og den sociale baggrund betyder, siger Knud Juel.

- Det kan man ikke gøre ret mange steder på nationalt niveau, som vi kan, så det er da med til at give nogle analyser, som vi ikke ville kunne få ellers. Så vi har større mulighed for mere præcist at belyse, hvad de sociale forhold betyder i Danmark, end man har mulighed for i mange andre lande.





Forskerbeskyttelsen og de faldende personstikprøver

De fleste data, som forskerne analyserer på, kommer fra administrative registre. Men for nogle analyser er dette ikke nok. Her må man supplere med spørgeskemaundersøgelser, hvor borgerne spørges om fakta eller holdninger. Danmarks Statistik udfører også sådanne undersøgelser, både for institutionen selv, for forskere og andre. Men svarprocenten har gennem en lang periode været faldende.

- Det er bare dalet kontinuert til i dag, hvor vi har omkring 60-65 pct. opnåelse, og Danmarks Statistik har stort set den højeste opnåelsesprocent, siger Peter Linde, som er kontorchef i Interviewservice.

Der kan være mange gisninger om, hvorfor deltagelsen i disse undersøgelser er faldet så drastisk. Mangel på tid, lyst og forståelse for vigtigheden er alle kvalificerede bud.

Forskerbeskyttelse medskyldig

En ting, der dog er svær at sætte spørgsmålstegn ved betydningen af, er den såkaldte forskerbeskyttelse. Den blev indført i 1995 og betød, at man som borger kunne få indsat en markering i CPR-registret om, at man ikke ønskede at

modtage henvendelser i forbindelse med statistiske og videnskabelige undersøgelser.

De første mange år var denne ordning dog ikke et problem. I 2001 var der nemlig blot registreret sølle 17 personer på denne liste, som på dette tidspunkt blev administreret af Indenrigsministeriet. Men da administrationen overgik til folkeregistrene ude i de enkelte kommuner kom et decideret boom.

Fra 17 til 800.000 personer

Da Folketinget efter indstilling fra Økonomi- og Erhvervsminister Margrethe Vestager i starten af 2014 valgte at afskaffe forskerbeskyttelsen, havde 800.000 borgere tilmeldt sig denne.

- Det gav nogle meget specielle pukler. Fx har der været en stor underrepræsentation på undersøgelser om studieønsker, siger Peter Linde og understreger, det ikke kun er Danmarks Statistik, der oplevede vanskeligheder ved forskerbeskyttelsen.

- Der er simpelthen bare flest undersøgelser, der direkte vedrører borgerne. Fx patienttilfredshedsundersøgelser og kommuner, som undersøger befolkningens til-

fredshed eller ønsker til dagsinstitutioner. Derudover er der måske ikke så mange, der interesserer sig for, hvad 28 forskere, der fylder 20 mio. kr. på finansloven mener, men det er vigtigt for samfundet at få borgerne til at vurdere patienttilfredsheden, og at unge mennesker tilkendegiver tilfredsheden med deres studievalg.



Datasikkerhed er hjerteblod

Danmark og de øvrige nordiske lande er i hele verden berømte for deres forskerliberale tilgang til national statistik. Dette kan kun lade sig gøre, hvis man har befolkningens tillid.

- Det er det vigtigste. For uden at folk har tillid til, at vi passer godt på data, og de ikke kommer ud og bliver brugt mod folk, så kan vi ikke have nogen statistik. Så det er simpelthen et kernepunkt for Danmarks Statistik. Det er hjerteblod, siger Lars Thygesen, som er afdelingsdirektør for bl.a. forskerservice og formand i Danmarks Statistiks datafortrolighedsudvalg.

Borgernes tillid har stor betydning for, at der kan laves så god statistik af så høj kvalitet som overhovedet muligt.

- Vi ved af erfaring, at dem, der ikke svarer i en undersøgelse, er anderledes end dem, som svarer. Og derfor får man skæve resultater, og man får underrepræsenteret visse sociale grupper i undersøgelse, og dermed bliver resultaterne ikke så brugbare. Fx vil for få arbejdsløse, for få med lave indkomster og for få med lave uddannelser svare, siger Lars Thygesen, som har kunnet se svarprocenten falde støt i Danmarks Statistiks undersøgelser siden 1970'erne.

Negative ringe i vandet

Der er flere bud på, hvorfor svarprocenten er faldende, bl.a. manglende tid og lyst. Men det smitter også af på Danmarks Statistik, når befolkningen hører historier i pressen om, at andre ellers troværdige institutioner ved en fejl kommer til at lække data med personnumre.

- Det undergraver også tilliden til Danmarks Statistik, selvom vi aldrig nogensinde har haft uheld med udslip af data. Det, at CPR-kontoret kommer til at lække personnumre (CPR-kontoret lagde ved en fejl 900.000 borgeres personnumre på nettet 2. juli 2014, red.), smitter af, selvom det ikke har noget med os at gøre. Vi kan se nu, der er en tendens til lavere svarprocent på vores undersøgelser i de sidste par måneder, og det, tror vi, skyldes, at man bliver mere mistænksom på grund af disse uheld andre steder, siger Lars Thygesen.

Størst grad af tillid i Europa

EU-Kommissionen foretager med mellemrum undersøgelser af borgernes tillid til deres lands statistikinstitution. Her indtager Danmarks Statistik hver gang en topplacering, men det betyder ikke, at man hviler på laurbærene.

- Det er jo dejligt, vi ligger sådan, men man kan ikke regne med, at det fortsætter, hvis man ikke hele tiden sørger for at passe meget på, og datasikkerhed er noget, der gælder hver dag, og som man skal tænke på hele tiden, siger Lars Thygesen, som mener, at den flotte placering både skyldes, at Danmarks Statistik generelt har et godt omdømme, og den generelle tillid der eksisterer i det danske samfund.

- Jeg har boet i Frankrig i nogle år, og der havde man grundlæggende mistillid til de offentlige myndigheder. Man mente, de var til for at genere borgerne. I Danmark mener man trods alt som hovedregel, at de offentlige myndigheder prøver at gøre noget godt for borgerne.

Krypteret og dynamisk

Den tillid er altafgørende for Danmarks Statistiks virke, og derfor er der også skrappe regler og hårde sanktioner, hvis nogen på et tidspunkt skulle overtræde reglerne.

- Hele forskerserviceordningen bygger på nogle sikkerhedskoncepter, som gør, at data ikke sådan kan smutte ud af systemet og blive brugt af andre. Forskerne arbejder på krypterede linjer og har nogle

sikkerhedstokens med dynamisk tildeling af adgangskode, siger Lars Thygesen.

Før man som forsker overhovedet får adgang til forskningsservice, skal både forskeren selv og dennes forskningsinstitution autoriseres. Her skal forskningslederen og forskeren skrive under på, at de vil følge reglerne.

Stikprøvekontrol hver dag

Med i alt 1.400 forskere tilknyttet forskerordningen er det umuligt for Danmarks Statistik at kontrollere hvert eneste output. Kontrollen foregår derimod – som med meget anden statistik - på stikprøvebasis hver eneste dag.

- Hvis noget ser mistænkeligt ud, så skrider vi ind. Det, forskerne må tage hjem, er resultater, estimater og tabeller, men de må ikke prøve at hjemtage enkeltoplysninger. Selvom de er afidentificerede, så er der jo en teoretisk risiko for, at man kan finde ud af, hvem det drejer sig om.

Overtrædelse medfører straf

Skulle en forsker på et tidspunkt i fremtiden komme til at overtræde reglerne, så er der to slags straf, man kan blive idømt. Den ene er, hvis man overtræder datalogivningen. Den slags medfører bøde eller i særlige tilfælde fængselsstraf.

- Men den mere nærliggende straf er, at man bliver udelukket fra at bruge forskerordningen for en periode eller måske for evigt, siger Lars Thygesen.

Det sker, hvis fx en forsker kommer til at overtræde reglerne og, uden at det nødvendigvis er af ond vilje, hjemtager ikke-identificerede personoplysninger.

Samarbejde med Datatilsynet

Danmarks Statistik og den danske forskningsservice bliver ofte rost i udlandet som en model, der giver en høj grad af frihed til forskerne. Dette skyldes ikke kun, at den danske statistikorganisation har været forskerliberal og fremsynet, men også et godt samarbejde med den kontrollerende myndighed, Datatilsynet, lyder vurderingen fra Lars Thygesen.

- Datatilsynet her i landet har forstået at tænke både på datasikkerheden og samtidig også på nytten af, at data bliver brugt til gode formål. Det er klart, hvis man sagde, at datasikkerheden skal være absolut, så ville man slet ikke kunne foretage nogen databehandling. Derfor gælder det for os, der arbejder for staten, om at afbalancere det, sådan man sørger for, der ikke er alvorlige risici, men at man samtidig kan gøre nyttige ting. Det er jo det, som hele forskningsservice går ud på.





Forskermaskinerne – teknikken bag

Der er mange menneskelige hjerner i sving, når der bedrives forskning. Men en altafgørende faktor er også forskermaskinerne, som er de maskiner, hvor forskerne laver deres datakørsler.

- Engang var der begrænsede mængder data, så man lavede ikke datakørsler på fuld population, de fik kun en stikprøve. Af plads- og processeringsmæssige årsager var man nødt til at begrænse forskningen til 10 procents populationer, siger systemkonsulent Henrik Philip Jørgensen.

Otto, Maria, Julie ...

Den første forskermaskine kom op at køre i 1988, og der er en tradition for at navngive sine servere. Den tradition videreførte Danmarks Statistik.

- På Københavns Universitet anvender man gamle gudenavne. Herinde har man knyttet noget af det op på personer, og da der kun var en forskermaskine, hed den Otto. Den var opkaldt efter Otto Andersen, siger it-chef Arne Brask Jørgensen.

Da der kom en *ny* maskine fra leverandøren Sun Microsystems, valgte man den kreative navngivning: Sunny!

Herefter blev forskermaskinerne opkaldt efter børn, børnebørn og elever: Oliver, Maria og Julie, og senere blev Sunny erstattet af Tonny. I dag – som følge af den efterhånden lange række af maskiner - er kreativitet blevet erstattet af pragmatisme.

- I dag kan man se ud fra navngivningen, hvad serverne laver. Det var lidt sværere tidligere, siger Henrik Philip Jørgensen.

Mindst 100 forskere hver dag

Antallet af forskere, der benytter forskningsservice, er vokset støt fra år til år, og i dag sidder mere end 100 forskere og arbejder med data hver eneste dag. På grund af stort pres på kapaciteten kan man som forskningsinstitution få lov at opstille sin egen server i Danmarks Statistik.

- Det er en service overfor forskerne, men en service de betaler for. Vi står for driften af maskinerne, tilser dem og sørger for, de er oppe at køre hver dag, siger Arne Brask Jørgensen.

I dag står der 80 eksterne servere i Danmarks Statistik i mange forskellige størrelser. Fra den allermindste, som er en laptop, til nogle store, kraftige servere. Serveren betaler forskningsinstitutionen selv for.

- Så får de deres forskningsdata flyttet derved på. Fordelen er, hvis de har behov for at køre nogle opgaver, der tager dage eller ugevis, så går det hurtigere. Eller hvis de vil have nogle programmer, der er helt specielle, så kan de installere dem på deres egen server, siger Arne Brask Jørgensen.

Hurtigere end hollænderne

De data, som forskerne sidder og arbejder med, stilles til rådighed af forskningsservice. Forskningsservice forbereder anonymiserede datasæt og stiller dem til rådighed. Forskerne kan af sikkerhedsmæssige hensyn ikke hente datasættene hjem, og de kan ikke overføre filer, men de kan få hjemsendt mellemresultater.

- Fra tid til anden bliver forskernes analyseresultater sendt hjem med e-mail til dem. På den måde laver vi også stikprøvekontroller, så vi kan se, hvad det er, de gerne vil have lov til at flytte hjem. De må ikke tage totale datasæt med hjem, siger Arne Brask Jørgensen.

Test hver anden time

For at sikre forskerne optimale arbejdsbetingelser, kører it-center en gang hver anden time et testjob.

- Man kan også måle cpu, diske og hukommelse, men det er bedre at have et program, som vi ved skal tage 50 sek. at køre igennem. Så kører vi sådan et job på en server hver anden time. Så tager vi tid og lægger den i en tabel. Så kan det godt være, den en sjælden gang sniger sig op på 100 sek., fordi en enkelt forsker presser maskinen. Derfor er sådan nogle jobs meget bedre at kigge på, når man skal kigge på ydelsen, for det er det, forskeren oplever. Han kan ikke bruge det til noget, at jeg siger, jamen hukommelsen har det fint. På den her måde holder vi løbende øje med situationen, siger Henrik Philip Jørgensen, som sammen med kollegerne i it-center månedligt afrapporterer til forskningsservice.

I afrapporteringen er en række parametre, som bedømmes efter en smileyordning: grøn, gul eller rød.

- Vi har nogle mål, og hvis der fx er under 10 pct. ledig hukommelse i mange dage, så er det ikke så godt. Så det her er vores guidelines til, hvornår vi skal investere, siger Arne Brask Jørgensen, som i de sidste mange måneder kun har set grønne, glade smileys. Men de skifter på et tidspunkt og bliver til gule, mellemfornøjede smileys.

- Det er netop det, der er meningen. Selvfølgelig ikke at de skifter, men at vi kan nå at gribe ind, når de begynder at blive gule. For så skal der gøres noget.





Lange tidsserier af høj kvalitet

For forskerne er et yndet mål at se på forskellige udviklinger i samfundet tilbage i tiden. Men når man fx skal sammenligne, hvordan danskernes nettoindkomst er i 2014 sammenlignet med i 1970'erne, så er der en masse faktorer, der spiller ind. Fx lovmæssige beslutninger.

For at overskueliggøre disse forskellige faktorer, har Danmarks Statistik udviklet en såkaldt "høj kvalitetsdokumentation", som er en overbygning på det fælles dokumentationssystem "TIMES". Her kan forskerne finde beskrivelser af, hvad hovedændringerne har været fra år til år på forskellige statistikområder.

- Der er til de mest brugte variable skrevet en grundig tekst, som i detaljer beskriver relevante ændringer over tid, vedlagt dokumenter og knyttet grafer til. Normalt vil der være en kontinuerlig udvikling, men nogen gange vil der være et hop, og så ved forskerne, at de skal passe på. Så det er et dokumentationsværktøj, som er guld værd for både forskerne og Danmarks Statistik, siger kon-torchef Ivan Thaulow, som fremhæver institutionens samarbejde med forskningsverdenen som den primære grund til, at dette system er udviklet.

- Systemet blev i høj grad udviklet på forskernes opfordring. Forskerne har i høj grad ære af dette, fordi de har skubbet på, og fordi de har kunnet skaffe finansiering til projektet via KOR (det Koordinerende Organ for Registerforskning, red.) og Forskningsministeriet.

Mange års arbejde forude

Det systematiske arbejde med at dokumentere tilbage i tiden startede i 2004.

- Det startede med, man tog en ordentlig klump variabler, bl.a. IDA, indkomster og demografi, og det var et kæmpe arbejde, som vi på daværende tidspunkt som institution ikke havde meget erfaring med. Derfor havde vi en stor bunke udokumenterede variabler de første år, men så fik vi ligesom hul på bylden. Det var også noget man skulle lære at acceptere i huset, at der kom en fuldmægtig fra forskningsservice og bad en kollega dokumentere ændringer langt tilbage i tiden, fortæller Ivan Thaulow.

I dag er der opstået et godt arbejdsflow omkring dokumentationen af tidsserier, og derfor har forskningsservice for nyligt rundet 600 dokumenterede variabler.

- Alt i alt har vi jo mange tusinde variabler, fordi vi har så mange registre, og nogle af disse registre er mere centrale end andre, så vi arbejder os fremad, men hvis vi forestiller os, at alle variabler skal ende sådan her, så er der mange års arbejde foran os. Vi klarer 20-30 variabler i gennemsnit hvert år. Der er selvfølgelig nogle, man kan komme hurtigt om ved, mens andre kan tage uger eller måneder.

Forskerner kvalitetssikrer

Det har indtil videre taget 10 år at nå dertil, hvor man er i dag, og hvert eneste år bliver Danmarks Statistik og forskningsverdenen enige om nye variabler, der skal dokumenteres.

- Vi tager dem, som bliver brugt mest, men vi har også en god samarbejdsform med forskerne, som stiller nogle forslag. Fx siger de i år socialpensioner, ældre-dokumentation og familie-indkomster, siger Ivan Thaulow.

Det tætte samarbejde med forskerne illustreres også af forskernes kontrol med det leverede arbejde.

- Forskerne kontrollerer det historiske arbejde, som vores folk har lavet, og det var

grænseoverskridende i starten. Folk nedlagde jo nærmest arbejdet herinde, fordi de ikke ville finde sig i, at forskerne skulle sidde og bestemme, siger Ivan Thaulow.

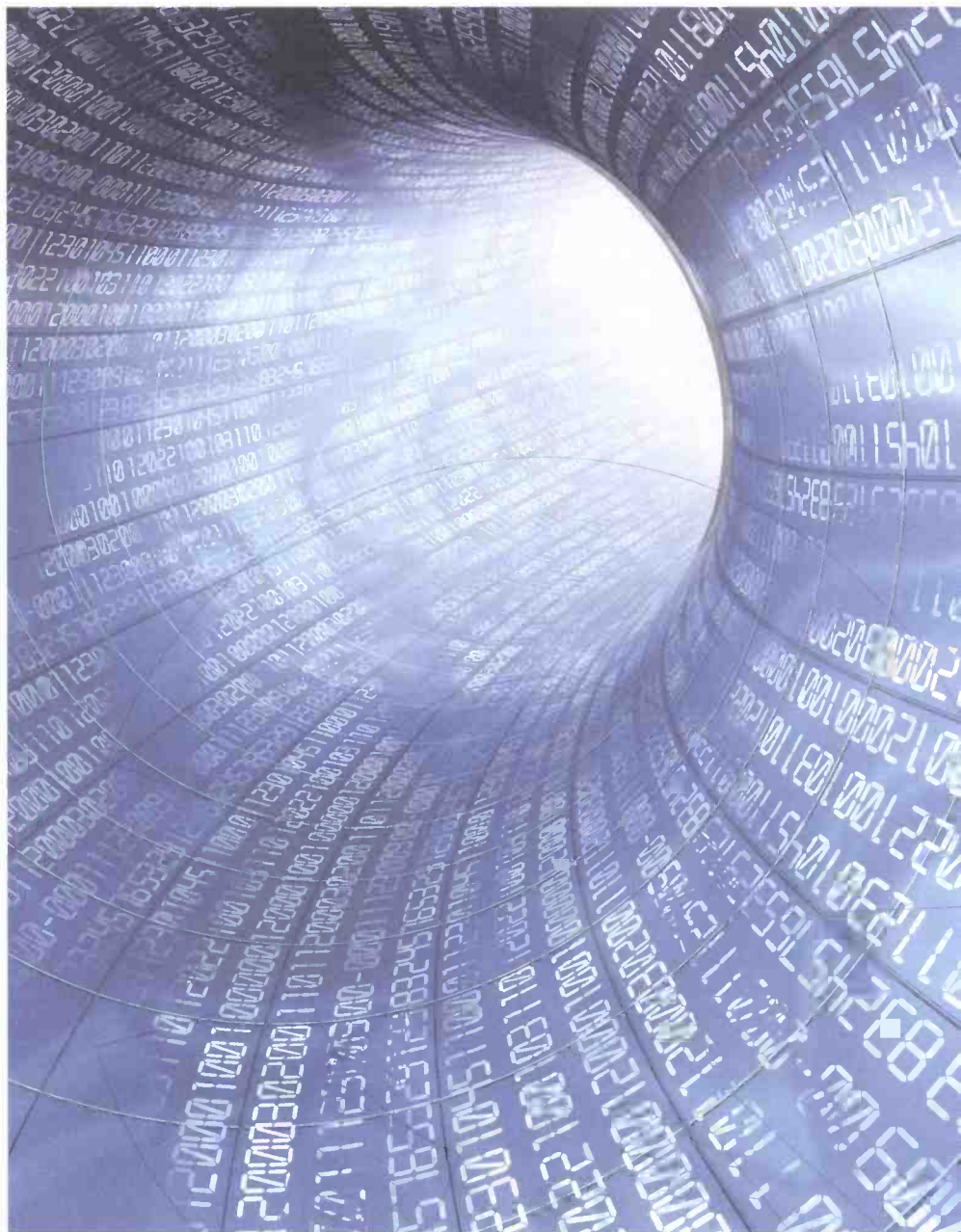
Nu har medarbejderne dog accepteret denne arbejdsgang, og når året er gået, indsender Danmarks Statistik sit arbejde, og så kommer forskerne med alle typer kommentarer. Så rettes det til og sendes ind til KOR, og først når de har sagt god for det leverede arbejde, udbetales pengene.

- Så det er lidt mere hardcore og noget af en kulturtilvænning, og det fungerer virkelig godt nu, siger Ivan Thaulow.

Udlandet op af stolene

De nordiske søsterorganisationer er ligesom Danmarks Statistik dygtige til at kvalitetsdokumentere, men med hensyn til dokumentation af lange tidsserier, så er Danmark stadig et skridt foran.

- Lige præcis det med at dokumentere tilbage i tiden, der kan vi stadig få de andre op af stolene af begejstring, for det er det, forskerne sukker efter, men også fordi der i Danmark er øremærket penge til dette. Det har de andre ikke fået igang endnu, men det skal nok komme, for det er jo helt åbenlyst en kæmpe gevinst, siger Ivan Thaulow.





Udlandet kigger med og kopierer den danske model

Forskningservice var i 1988 unikt på verdensplan. Siden da har den danske statistikinstitution ikke set sig tilbage og betragtes den dag sammen med sine nordiske søsterorganisationer og Nederlandene som internationalt førende.

- Det er jo en af Danmarks styrkepositioner, at vi har alle de her registre, og at vi kan udnytte dem til forskningsformål, siger Elsebeth Lynge, professor ved Center for Epidemiologi og Screening på Københavns Universitet, som gennem årene har hørt meget positivt, når hun har talt med udenlandske kolleger om forskernes muligheder for at få data.

- Man er i hele verden meget bevidst om, at Norden har en speciel position. Man har i Norden dels registrene, men man har også befolkningens tillid til de offentlige myndigheder, og så har man nogle effektive offentlige myndigheder, så registrene også passer på virkeligheden. Men man kan så sige, at det danske statistikbureau var hurtigere til at lukke op for de her forskningsmuligheder end de andre nordiske lande.

Udlandet stærk inspireret

Den vurdering er tidligere rigsstatistiker Jan Plovsing enig i.

- Forskningservice har betydet meget, og man har også kigget på det udefra. Vi har været ude i verden for at fortælle om systemet. Svenskerne har sidenhen udarbejdet noget, der helt klart – og det sagde de også bagefter – er inspireret af vores system. Vi har haft en del andre landes statistikinstitutioner til at komme i min tid for at se på og få fortalt, hvordan det her system fungerer.

Det er dog ikke kun nationale statistikinstitutioner, der har været på besøg i Danmark for at suge inspiration. Blandt andet har OECD og Eurostat, som er EU's officielle statistiske kontor, været på inspirationsstur til Danmarks Statistik.

Specielle registermuligheder

Jan Plovsing fremhæver ligesom Elsebeth Lynge de danske registre som noget helt særligt i de danske muligheder for at bedrive forskning på højeste niveau.

- Lad os bare tage et land som USA. Her har man slet ikke et register over hele befolkningen, som man kan kombinere med alt muligt om den enkelte borger. Det gør, at fx amerikanske forskere synes, det er så fantastisk.

I juni 2014 udkom i Sverige en regeringsrapport om registerforskning. Op til denne udgivelse havde svenskerne været i Danmark for at få god inspiration til, hvordan man kan organisere det, og de var imponerede, fortæller Elsebeth Lynge, som selv deltog.

- De var en hel dag ude og se forskningservice. Det er klart, at vi har nogle muligheder, som andre lande ikke har, og så har Danmark været lidt hurtigere til at udnytte disse muligheder. Det er en kombination af, at Danmarks Statistik med Otto Andersen indførte en mere forskerliberal holdning, men jeg tror også, man ikke skal underkende, at Datatilsynet i Danmark også har haft en mere forskervenlig tilgang end Datatilsynet i de andre nordiske lande, lyder vurderingen fra Elsebeth Lynge.

Tættere nordisk samarbejde

Netop de nordiske nabolande kommer sandsynligvis til at spille en vigtig rolle i fremtiden i forskningservice. For der arbejdes lige nu et tættere internordisk samarbejde til fordel for forskerne.

- Vi har en dugfrisk rapport, som er lavet af en arbejdsgruppe, der har fået til opgave at beskrive, hvilken type af nordisk

fællesordning, vi kan lave for at gøre det lettere for forskerne at lave tværnordiske projekter. Det hele udspringer fra forskernes ønsker, siger Ivan Thaulow, som har været formand for arbejdsgruppen.

Hele forslaget bunder i det faktum, at forskerne bliver stadig mere internationale, og derfor er behovet for at få fat på data på tværs af landegrænser også stigende.

- Vi har med vores datasikkerhed altid været blandt de bedste i verden, fordi vi har haft nogle tykke sikkerhedsmure, men har tilladt forskerne at komme igennem disse mure under nogle bestemte ordninger. Nu er de nordiske lande ved at komme med, så de har mure, der er ligeså tykke som vores. Altså er sikkerheden i dag ligeså højt prioriteret her som i Danmark. Derfor tør vi godt arbejde tættere sammen med dem om forskning i fællesnordiske data.

Den store drøm

- Drømmen er selvfølgelig for de nordiske forskere, at man pludselig får en befolkning, som består af mere end 25 millioner mennesker istedet for fx kun knap 6 mio. i Danmark, siger Ivan Thaulow.

Der kommer dog stadig til at være vetoret for hvert enkelt land i hver enkelt sag.

- Alle lande siger, at vurderingen af et projekt skal være fra case til case. Forskerne skal fortælle, hvad de gerne vil forske i, og så skal hvert enkelt land fortsat have retten til at sige nej. Dermed bevarer vi den

nationale selvbestemmelse. Vi vil sikre, at nordiske forskningsprojekter, der får adgang til danske data, overholder danske sikkerhedsbetingelser, siger Ivan Thaulow.





Forskningservice gennem 25 år

Udgivet: November 2014

Tryk: Rosendahls-Schultz

Fotos: Imageselect
Colourbox
Polfoto
Danmarks Statistik

Layout: Leila L. Hjulgaard

Forfatter: Henrik Alexandersen

Redaktion: Lars Thygesen
Ivan Thaulow
Carsten Zangenberg