

K O M M E N T A R

af Anne Sofie Fink Kjeldgaard

Bevaring af forskningsdata – hvorfor og hvordan?

Artiklen præsenterer argumenter for, at forskningsdata arkiveres og formidles til brug for nye forskningsformål. Heri ligger en opfordring til, at arkivering integreres i arbejdet med empiriske data, så sikring og udnyttelse af data fremstår som del af forskningsprocessen. Artiklen afsluttes med nogle visioner for arkivering og dataformidling.

Når et forskningsprojekt indsamler empiriske data, skabes der en resurse ikke alene for det enkelte projekt, men også potentielt for andre fremtidige forskningsprojekter enten af forskeren selv eller af andre forskere og/eller studerende. Den form for genbrug af eksisterende data til nye analyser betegnes sekundæranalyse (fx Hyman 1972, Hakim 1982). Når sekundæranalyse er en vigtig forskningspraksis, skyldes det, at den primære analyse af data vil være begrænset til at berøre et udsnit af det indsamlede datamateriale, hvorfor genanvendelse af data kan medvirke til en mere bredspektret anvendelse (Rasmussen 2000).

Formidling af data i Danmark

Tilbage i 1960'erne opstår tankerne om, at indsamlede kvantitative forskningsdata på elektronisk form bør bevares som dokumentation af den udførte forskning og som potentiale for nye analyser. I Danmark manifesterer disse tanker sig i etablering af et dataarkiv for samfundsbeskrivende data i 1973. Arkivet får navnet Dansk Data Arkiv (DDA) (www.dda.dk) og bliver på den måde ikke snævert rettet mod den samfundsvidenskabelige forskning. Det er dog i første omgang samfundsvidenskaben, der tilbydes service fra DDA, men også historie¹ og sundhedsvidenskab² kommer til. Inden for de senere år er der kommet to yderligere datadistributører til den danske datainfrastruktur. Det drejer sig om Surveybanken ved Ålborg Universitet (www.surveybank.aau.dk) og Sociologisk Databank ved Københavns Universitet (databank.soc.ku.dk).

Hvorfor arkivere forskningsdata?

Empirisk forskning inden for samfundsvidenskab er kostbar, og en væsentlig del af omkostningerne går til indsamling af data (Rasmussen 2000). Det betyder, at man alene ud fra en samfundsøkonomisk vinkel kan argumentere for arkivering og sekundæranalyse, fordi det betyder mere forskning for de samme økonomiske midler. Betraget ud fra en samfundshistorisk vinkel kan der argumenteres for, at samfundsvidenskabelige data udgør vigtig dokumentation af sin tid. Vendes perspektivet til den almene borger, kan genbrug af data betyde, at kravene, der lægges på befolkningen om at agere respondenter i diverse samfundsvidenskabelige undersøgelser, begrænses (Hakim 1982).

Argumentet for dataarkivering for den enkelte forsker er tosidet. For det første indebærer arkivering i DDA, at datamaterialet gennemgår en oparbejdningsproces, der giver materialet en standardiseret form. Datamaterialet får i kraft af oparbejdning højere kvalitet end oprindeligt³ og kan betragtes som "uforældeligt" uanset ændringer i programmer og/eller formater i fremtiden. Det sikres desuden, at den nødvendige dokumentation er til stede for, at materialet kan forstås og eventuelt anvendes til sekundæranalyse. Oparbejdningsservicen er unik for DDA. For det andet betyder arkivering, at primærforskeren ikke selv skal være ansvarlig for formidling af datamaterialet. Den opgave varetages af dataarkivet eller -banken i henhold til forskerens ønsker om adgangsbegrænsning. Dataformidling er afhængig af datamaterialets kvalitet – jo bedre datakvalitet, desto lettere forløber formidlingen af datamaterialer (Gregory et al. 2009).

Nogle af de offentlige forskningsråd bl.a. Det Frie Forskningsråd for Samfund og Erhverv stiller krav om, at støttede projekter skal arkivere indsamlede data i DDA. Forskeren forpligter sig altså ved modtagelse af bevillingen på dataarkivering. For at sikre den bedst mulige sikring og udnyttelse af forskningen og midler forbrugt hertil kunne også andre fonde og -råd samt tidsskrifter overveje at stille lignende krav til datas tilgængelighed (Udvalget vedr. sundhedsvidenskabelige data 2000).

Hvis datamaterialer indeholder følsomme persondata som fx navne, adresser eller personnumre, forlanger Datatilsynet til gengæld, at data slettes ved projektets ophør. Måden, data kan bevares på, er ved overførsel til DDA, fordi arkivering i DDA af Datatilsynet sidestilles med sletning. Al udlevering af det arkiverede datamateriale med følsomme personoplysninger fordrer Datatilsynets tilladelse (for mere information se Daasnes 2008).

Formidling af data

I forhold til dataarkivets formidlingsaktivitet opridser Hakim (1982) en række forhold, der kan være motiv for forskere og studerende til at søge adgang til arkivdata. Hun lægger især vægt på de besparende aspekter, der følger med anvendelse af sekundæranalyse: 1) modvilje mod at binde store økonomiske og tidsmæssige ressourcer i dataindsamling, 2) begrænsede økonomiske og/

eller tidsmæssige ressourcer til rådighed hos forskere og studerende, 3) manglende motivation til at gennemføre dataindsamlingsprojekter samt 4) kobling af empiri til et overvejende teoretisk forskningsprojekt. Sætter man også fokus på de tilføje og vidererækkende aspekter i anvendelse af arkiverede data, vil det handle om at udnytte de forskningsmæssige muligheder, der ligger i at have flere datamaterialer let tilgængelige. Det kunne eksempelvis være udnyttelse gennem analyser af enkelte forhold på tværs af flere undersøgelser, at gennemføre analyse med et andet teoretisk udgangspunkt end primærforskerens, at sammenligne flere undersøgelser om samme emne gennemført på forskellige tidspunkter eller, at eksisterende datamaterialer analyseres i forhold til senere opståede samfundsmæssige debatter (Kjeldgaard et al. 2008).

Hvordan arkiveres data?

Arkivering af et datamateriale – hvad enten det er et kvantitativt eller kvalitativt materiale – indebærer aflevering af et data og datadokumentation til DDA. Datadokumentation er den information, der gør et datamateriale forståeligt og anvendeligt. Det drejer sig om information om primærundersøgelser, dataindsamlingstidspunkt, udvælgelse af respondenter, bortfald, oversættelse af talkoderne i spørgeskemaet mm. Datadokumentation består som minimum af DDA's afleveringsskema og undersøgelsens spørgeskema eller -spørgeguide.

Visioner for fremtiden

Mange spørgeskemaundersøgelser foretaget af forskere i Danmark formidles af DDA og/eller databanker. Hermed bevares data, en samfundsmæssig virkelighed dokumenteres, og materialet stilles til rådighed for nye forsknings- og studieformål i det omfang, den afleverende forsker tillader det. Her ligger et enormt materiale for forskning og undervisning.

Internationalt samarbejde mellem dataarkiver i forbindelse med udvikling af arkiveringsstandarden DDI3 – *Data Documentation Initiative* – betyder, at der kan forventes væsentlige forbedringer af, hvordan datamaterialer arkiveres og gøres tilgængelige i fremtiden. Det handler primært om, at forskerne tilbydes *software*-produkter til løbende dokumentation af dataindsamlingsprocessen, og at datadistributører kan "linke" relaterede datamaterialer, så fx muligheden for sammenligning af data/variable på tværs af datamaterialer understøttes. I løbet af 2010 kan forventes introduktion af et fælles europæisk søgekatalog baseret på DDI3-standarden.

Det er vigtigt, at der sikres den lettest mulige adgang til de arkiverede data, så datas potentiale udnyttes optimalt. Det betyder blandt andet, at der løbende må arbejdes på at forbedre formidling af data og dokumentation via internettet. Ideelt set skal en service som eksempelvis DDA's imidlertid ikke være en enkeltstående resurse. Dataarkivering og -formidling skal kobles til andre former for forskningsinformation og -publikation eksempelvis i form af

emnebaserede web-steder, der tilbyder mange forskellige resurser vedr. data, publikationer, netværk, internationale resurser, lovgivning etc.

Jeg håber, artiklen her med sin argumentation for både det samfundsmæssigt og det individuelt nyttige i dataarkivering og -formidling har motiveret til, at flere afleverer data og udnytter eksisterende data i forskning og undervisning.

Kontaktinformation:

Anne Sofie Fink Kjeldgaard, asf@dda.dk

Dansk Data Arkiv (DDA)

Islandsgade 10

5000 Odense C

Noter

1. DDA har fra sin start modtaget indtastede kvantitative opgørelse fra historisk forskning. Hertil kommer, at DDA har stået i spidsen for kildeindtastningsprojektet (KIP), der med frivilliges hjælp har fået indtastet folketællinger og kirkebogsoplysninger til brug for demografisk forskning og slægtsforskning.
2. DDA har siden 1996 arkiveret og formidlet sundhedsvidenskabelige data først i enheden ERAS – Enheden for registrering og arkivering af sundhedsvidenskabelige data – og siden 2004 i DDA Sundhed.
3. Eksempelvis sikres det, at alle variable er dokumenteret, hvis ikke søges det at indhente informationen fra forskeren, der har afleveret materialet, og det checkes, om filtreringen af data er foretaget korrekt, hvis ikke indsættes en kommentar om det. Det skal understreges, at DDA ikke foretager nogle former for ændringer i datamaterialet.

Litteratur

- Corti, Louise, Witzel, Andreas and Bishop, Libby (eds.) 2005: "Secondary analysis of qualitative data", *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(1), qualitative-research.net/fqs/fqs-e/inhalt1-05-e.htm
- Daasnes, Camilla 2008: "Persondataloven – regler og praksis for god databehandlingskik". *Metode & Data*, nr. 94, 2008.
- Fink, Anne Sofie 2000: "Kvalitative data i Dansk Data Arkiv". *Metode & Data*, nr. 83, 2000:2.
- Gregory, Arofan, Pascal Heus & Jostein Ryssevik 2009: "Metadata". Working Paper No. 57. German Council for Social and Economic Data (RatSWD).
- Hakim, Cathrine 1982: *Secondary Analysis in Social Research*. London: George Allen & Unwin.
- Hyman, Herbert H. 1972: *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentials*. New York: John Wiley & Sons, Inc.
- Kjeldgaard, Anne Sofie Fink, Lone Bredahl Jensen & Bodil Stenvig 2008: "Genbrug af forskningsdata – Overset tilbud eller anvendt mulighed". *Konferenceudgivelse: Symposium i anvendt statistik 2008*.
- Rasmussen, Karsten Boye 2000: *Datadokumentation Metadata for samfundsvidenskabelige undersøgelser*. Odense: Odense Universitetsforlag.
- Udvalget vedr. sundhedsvidenskabelige data 2000: *Opbevaring og arkivering af sundhedsvidenskabelige forskningsdata – et debatoplæg*. Statens Sundhedsvidenskabelige Forskningsråd.