

Dokumentasjonsguide for arkivering av forskningsdata

Ved arkivering av forskningsdata bør datasettet dokumenteres på tre nivå: Prosjektnivå, datasettnivå og variabelnivå. På **prosjektnivå** for å beskrive hensikten med prosjektet, hvem som er ansvarlige for gjennomføringen og de metodene som er tatt i bruk. På **datasettnivå** bør det gis en oversikt over de ulike filene som følger med i overføringen slik at man vet hva de inneholder og hvordan filene er relatert til hverandre. På **variabelnivå** må dokumentasjonen beskrive selve dataene på en slik måte at de gir mening for andre. Det bør gis en oversikt over alle variabler og koder og en forklaring av dem.

Prosjektnivå:

Tittel

Kontekst – prosjektbeskrivelse (hensikt, problemstilling/hypotese, eksisterende forskning etc), navn på person og/eller institusjon som har ansvar for prosjekt, hvem som har gjennomført datainnsamling og hvem som skal ha de intellektuelle rettighetene til datamaterialet. Eventuelle finansieringskilder og samarbeidspartnere.

Metode for datainnsamling – sampling, innsamlingsprosedyrer, skalaer, programvarer og utstyr tatt i bruk, instrumenter, geografisk dekning, tidsperioder, datatype, og bruk av sekundærdata

Tilgang – informasjon om hvem som skal få tilgang til data fra prosjektet og når dataene skal gjøres tilgjengelig.

Eksempel:

Tittel	Samordnet levekårsundersøkelse 2009 - Tversnitt Tema: Arbeidsmiljø
Ansvarlig forfatter	Statistisk sentralbyrå
Opphavsrett	Copyright © 2009 Statistisk sentralbyrå
Sammendrag	Statistisk sentralbyrås Samordnet levekårsundersøkelse - tverrsnittundersøkelsen (temaroterende) har som hovedmål at den over tid skal dekke alle viktige levekårsområder. I tverrsnittdelen har temaene rotert som følgende: Arbeidsforhold i 1996, 2000, 2003 og 2006; Boforhold, nærmiljø og fritid i 1997, 2001 og 2004; Helse i 1998, 2002, 2005 og 2008. Hovedtema for Levekårsundersøkelsen 2009 er arbeidsmiljø.
Innsamlingsdato	Start: 2009-06-11 Slutt: 2010-01-09
Tidsperiode dekket	2009-2010
Land	Norge
Geografisk dekning	Land
Tidsmetode	Tverrsnittstudie
Datatype	Intervjudata (spørreundersøkelse)
Datainnsamler	Statistisk sentralbyrå (SSB)
Utvalg	Personer i Norge, 16 år og eldre. Bruttoutvalg: 20460 personer Nettoutvalg: 12255 personer Svarprosent: 59,9%
Tilgang	Data er tilgjengelig for forskningsformål etter eksplisitt søknad etter dato: 2015-12-31.

Datasettnivå:

Struktur – navn og format på datafiler, beskrivelse av ulike cases (tid, sted, utvalg) og innhold i fil, og forholdet mellom de ulike filene.

Versjonskontroll – versjonsnummer og dato for opprettelse av fil. Navn på ansvarlig person.

Anonymisering – Dokumentasjon på eventuell tillatelse til oppbevaring av data/forklaring på hvordan datasettet er blitt anonymisert (slettet navn/adresser, erstattet fødselsdato med aldersgrupper, generalisert betydningen av detaljert tekst – som for eksempel å erstatte en doktors spesifikke ekspertiseområde med en bredere medisinsk retning, bruk av pseudonymer, eliminering av outliere etc).

Eksempel:

Format	Navn	Størrelse kB	Versjon	Dato opprettet	Dato sist endret	Beskrivelse
PDF	Datasett_dokumentasjon.pdf	1403	02.01	01/06/2007	20/01/2008	Dokumentasjon om datasettet (liste over variabler, avledede variabler)
SPSS	Video_observasjoner_00_05	625	00.05	04/07/2007	25/11/2007	Resultatet av 120 videoobservasjoner utført i 5 barnehager i Essex, juni 2007

Variabelnivå:

(Gjelder for strukturerte/kvantitative data):

Navn, labels og beskrivelse av variabler, registreringer og deres verdier. Spørsmålsformuleringer ved spørreskjemaundersøkelser. (spørreundersøkelser/kodebøker)

Forklaring på koder, klassifiseringsordninger, terminologi og forkortelser tatt i bruk. Koder og grunner for missing values, samt beskrivelse av konstruerte variabler.

Eksempel:

Datasett: Undersøkelse om arbeid, livsstil og helse, 2010					
Variabelgruppe	Navn	Label	Beskrivelse	Verdier	Forklaringer
Arbeid og sysselsetting	v009	spm9 Fast eller midlertidig ansatt	Spørsmålstekst: Er du fast eller midlertidig ansatt i jobben du har nå?	1 – Har fast stilling 2 – Er midlertidig ansatt/vikar/engasjement e.l. 8 – vil ikke svare 9 – Vet ikke	
	v010	spm10 Avtalte arbeidstid pr. uke	Spørsmålstekst: Hva er din avtalte arbeidstid pr. uke? <i>Regn ikke med overtid eller fravær på grunn av sykdom e.l. Oppgi arbeidstiden i hele timer.</i>		

Helse	v095	Spm33 Vurdering av egen helse	Spørsmålstekst: Hvordan vurderer du din egen helse sånn i sin alminnelighet?	1 – Svært god 2 – God 3 – Verken god eller dårlig 4 – Dårlig 5 – Svært dårlig 8 – Vil ikke svare 9 – Vet ikke	
	v096	Spm34 Noen som står IO nær som IO kan snakke med	Spørsmålstekst: Bortsett fra medlemmer av din egen familie, har du noen som står deg nær, og som du kan snakke fortrolig med?	1 – Ja 2 – Nei 3 – Vet ikke 8 – Vil ikke svare 9 – Vet ikke	IO = Intervjuobjekt